第一卷

数学名著译丛

数学

它的内容，方法和意义

〔俄〕A. D. 亚历山大洛夫 等 著

科学出版社
www.sciencep.com

VISIT…

LANZAROTE Caliente.com

2R19/26

内 容 简 介

本书是苏联数学界普及数学知识的一部名著．全书共二十章，分三卷出版．书中不仅综述了现代数学的各分支学科，还综述了数学的哲学、历史发展及其在物理和工程技术方面的应用．本书每章都由苏联第一流学者撰写，有很高的学术水平．同时，本书写得深入浅出，只要具备高中的数学知识，就可以理解．在努力实现四个现代化的今天，这部书必将帮助我国广大读者扩大眼界，找到解决问题的数学工具．本卷包括四章，内容是数学概观、数学分析、解析几何、代数．

本书可供大学数学系师生、中学教师、具有高中以上文化程度的学生、有关方面的科技人员及数学爱好者阅读．

А. Д. Александров

МАТЕМАТИКА, ЕЁ СОДЕРЖАНИЕ, МЕТОДЫ И ЗНАЧЕНИЕ/ТОМ 1/

Издательство Академии Наук СССР, 1956

数 学

——它的内容、方法和意义——

〔苏〕А. Д. 亚历山大洛夫 等著

孙小礼 赵孟养 裘光明 严士健 译

关肇直 秦元勋 校

科学出版社出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1958 年 8 月第一版 开本：850×1168 1/32

1986 年 6 月第六次印刷 印张：10 1/4

印数：85,221—90,420 字数：269,000

统一书号：13031·2479

本社书号：3405·13—1

定价：2.30 元

重印说明

数学是研究现实世界的空间形式和数量关系的科学，它在自然科学和技术的各个部门有着广泛的应用，对人类认识自然和改造自然起着重要的作用．近年来，随着科学技术的迅速发展，数学的研究范围不断地扩大，内容日益丰富，已发展成为分支众多的庞大系统．由于数学的理论具有高度抽象的形式，它的许多较新的分支不易为非专业的人员所理解．为使读者能在较短的时间内用较少的精力获得关于数学的起源、现状及发展的基本知识，苏联的数学家们编写了这部著作．

本书包括数学的二十个分支，分别由苏联第一流数学家撰写，对于数学的各个分支、内容及方法，它们赖以建立的基础以及它们的发展过程作了简明而系统的论述．书中并阐述了数学的哲学，数学的发展对科学文化的影响，以及数学在物理和工程技术方面的应用．各个学科的科学家及工程技术人员都会在本书中找到有用的内容，而对于一些年青的数学家来说，本书也可开扩其眼界，使他们了解数学的全貌．

本书叙述由浅入深，内容融汇贯通，而其所要求的预备知识并不很多，只要具备高中的数学知识就可理解本书的大部分内容．

本书中译本于五十年代末出版，受到广泛的欢迎，至今仍有许多读者要求重印．因纸型已毁，现据原译本做了一些必要的修改，重新排印．

当然，本书问世已有二十余年．在这期间，数学发展突飞猛进，有许多新的内容，新的概念，本书未能包括．但是，对于了解数学的基本内容和历史发展，本书仍不失为一部优秀的参考读物．至于想要了解数学最新发展的读者，可参阅其他有关著述．

科学出版社

第 二 卷

第 五 章　常微分方程(И. Г. 彼得罗夫斯基著)
第 六 章　偏微分方程(С. Л. 索伯列夫著)
第 七 章　曲线和曲面(А. Д. 亚历山大洛夫著)
第 八 章　变分法(В. И. 克雷洛夫著)
第 九 章　复变函数(М. В. 凯尔迪什著)
第 十 章　质数(К. К. 马尔德尔扎尼吉维里著)
第十一章　概率论(А. Н. 柯尔莫果洛夫著)
第十二章　函数逼近法(С. М. 尼阔尔斯基著)
第十三章　近似方法与计算技术(В. И. 克雷洛夫著)
第十四章　电子计算机(С. А. 勒贝杰夫著)

第 三 卷

第十五章　实变函数(С. Б. 斯捷奇金著)
第十六章　线性代数(Д. К. 法德杰也夫著)
第十七章　抽象空间(А. Д. 亚历山大洛夫著)
第十八章　拓扑学(П. С. 亚历山大洛夫著)
第十九章　泛函分析(И. М. 盖尔芳特著)
第二十章　群及其他代数系统(А. И. 马尔采夫著)

原　　序

数学，由于实际的需要在古代便已经产生了，现在发展成为分支众多的庞大系统．数学正如其他科学一样，反映了物质实际的规律，并成为理解自然和征服自然的有力武器．但由于数学本身的高度抽象性，致使它的新的部门比较难为非专业的人所理解．正因为数学的这种抽象特征，所以还在古代便产生了认为数学与物质实际无关的唯心概念．

在编写这本书时，作者们是从这样的共同愿望出发，即要向苏联知识界的相当广大的阶层介绍每个数学分支的内容与方法，它的物质基础及发展道路．

读者只要具备中等学校数学课程的知识就能阅读本书；但三卷中每卷材料的难易程度是不一致的．要想初步认识高等数学的原理，可读前面几章；但要全部理解以后各章，则需要参考相应的教科书．对全书而言，则基本上只有那些在运用数学分析方法（微分法与积分法）已有某些经验的读者才容易理解．对于这类读者——自然科学与工程专业界人士及数学教师——引导他们熟悉更新的数学分支的那些章节是特别重要的．

自然，要在一部书里概括数学研究的丰富内容（即使是它几个主要方向的），是不可能的，因此在选材方面就必须有某些自由．但总的说来，这部书应当能使读者对近代数学的情况及其发生与整个发展的前景大致有一个概念．因此在一定程度上也考虑到那些已知道书中所用的事实材料的基本部分的人．这本书当能帮助我们的某些青年数学工作者消除他们有时常有的某些眼界的狭隘性．

本书各章由不同的作者写成，作者的姓名分载于目录中．但作为一部完整的著作来说，则是一个集体劳动的产物．它的总的

计划、材料的选择、各章文稿的内容，都经过集体讨论，并在热烈地交换意见的基础上加以改善. 苏联很多城市的数学家在有关的讨论会上对本书的初稿发表了宝贵的意见，这些意见和建议，作者们都曾加以考虑.

一部分作者也直接参加了其他各章的最后定稿工作：第二章的绪论部分基本上是 Б. Н. 杰龙涅写的；Д. К. 法捷耶夫积极地参加了第四章及第二十章的编写工作.

除了各章的作者外，还有以下同志参加了工作：Л. В. 康托罗维奇写了第十四章第四节，О. А. 拉得任斯卡雅写了第六章第六节，А. Г. 波斯特尼可夫写了第十章第五节，О. А. 俄列尼克参加了第五章文稿的编写工作，Ю. В. 普罗霍罗夫参加了第十一章文稿的最后校订工作.

В. А. 扎尔加列尔写了第一、二、七及十七等章的若干节. 文稿的最后校订是由 В. А. 扎尔加列尔及 В. С. 维金斯基在 Т. В. 洛果兹金娜及 А. П. 列奥诺娃的参与下完成的.

插图的主要部分是由 Е. П. 谢金绘制的.

编辑委员会

秦元勋 译

目　录

第　一　卷

原序 ……………………………………………………………………………v

第一章　数学概观(A. Д. 亚历山大洛夫著)……………………………1

§1. 数学的特点 ……………………………………………………1

§2. 算术 ……………………………………………………………7

§3. 几何……………………………………………………………18

§4. 算术和几何……………………………………………………23

§5. 初等数学时代…………………………………………………33

§6. 变量的数学……………………………………………………40

§7. 现代数学………………………………………………………53

§8. 数学的本质……………………………………………………61

§9. 数学发展的规律性……………………………………………71

第二章　数学分析(M. A. 拉夫伦捷夫、C. M. 尼阔尔斯基合著) ……………………………………………………………81

§1. 绪论……………………………………………………………81

§2. 函数……………………………………………………………89

§3. 极限……………………………………………………………97

§4. 连续函数 ……………………………………………………104

§5. 导数 …………………………………………………………109

§6. 微分的法则 …………………………………………………118

§7. 极大与极小, 函数图形的研究 ……………………………124

§8. 函数的增量与微分 …………………………………………134

§9. 泰勒公式 ……………………………………………………140

§10. 积分 …………………………………………………………145

§11. 不定积分, 积分的技术 ……………………………………154

§12. 多元函数 ……………………………………………………159

§13. 积分概念的推广 ……………………………………………173

§14. 级数 …………………………………………………………181

)4/1159/08

第三章　解析几何 (Б. Н. 狄隆涅著) ……………………………196
§ 1. 绪论 ……………………………………………………………196
§ 2. 笛卡儿的两个基本观念 ………………………………………197
§ 3. 一些最简单的问题 ……………………………………………199
§ 4. 由一次和二次方程所表示的曲线的研究 ……………………201
§ 5. 解三次和四次代数方程的笛卡儿方法 ………………………203
§ 6. 牛顿关于直径的普遍理论 ……………………………………206
§ 7. 椭圆、双曲线和抛物线……………………………………………208
§ 8. 把一般的二次方程化成标准形状 ……………………………220
§ 9. 用三个数规定力、速度和加速度. 向量理论……………………226
§ 10. 空间解析几何. 空间中的曲面的方程和曲线的方程 ………232
§ 11. 仿射变换和正交变换 …………………………………………240
§ 12. 不变量理论 ……………………………………………………251
§ 13. 射影几何 ………………………………………………………255
§ 14. 罗仑兹变换 ……………………………………………………262
结束语 ……………………………………………………………………270
第四章　代数(代数方程的理论) (Б. Н. 狄隆涅著) ……………273
§ 1. 绪论 ……………………………………………………………273
§ 2. 方程的代数解 …………………………………………………277
§ 3. 代数基本定理 …………………………………………………292
§ 4. 多项式的根在复平面上的分布的研究 ………………………302
§ 5. 根的近似计算法 ………………………………………………313

第一章 数学概观

对于任何一门科学的正确概念，都不能从有关这门科学的片断知识中形成，尽管这些片断知识足够广泛．还需要对这门科学的整体有正确的观点，需要了解这门科学的本质．本章的目的就是给出关于数学的本质的一般概念．为了这个目的，没有很大必要去详细考察新的数学理论，因为这门科学的历史和初等数学就已经提供了足够的根据来作出一般的结论．

§1．数学的特点

1．甚至对数学只有很肤浅的知识就能容易地察觉到数学的这些特征：第一是它的抽象性，第二是精确性，或者更好地说是逻辑的严格性以及它的结论的确定性，最后是它的应用的极端广泛．

抽象性在简单的计算中就已经表现出来．我们运用抽象的数字，却并不打算每次都把它们同具体的对象联系起来．我们在学校中学的是抽象的乘法表——总是数字的乘法表，而不是男孩的数目乘上苹果的数目，或者苹果的数目乘上苹果的价钱等等．

同样地在几何中研究的，例如，是直线，而不是拉紧了的绳子，并且在几何线的概念中舍弃了所有性质，只留下在一定方向上的伸长．总之，关于几何图形的概念是舍弃了现实对象的所有性质只留下其空间形式和大小的结果．

全部数学都具有这种抽象的特征．关于整数的概念和关于几何图形的概念——这只是一些最原始的数学概念．之后才是其他许多达到象复数、函数、积分、微分、泛函、n维甚至无限维空间等等这样抽象程度的概念．这些概念的抽象化好象是一个高于一个，一直高到这样的抽象程度，以致看上去已经失去了同生活的一切

联系，以致"凡夫俗子"除了感到"莫名其妙"以外什么也不能理解.

事实上情形当然不是这样. 虽说 n 维空间的概念的确非常抽象，但它却有完全现实的内容，要了解这内容并不那么困难. 在这本书里将要特别强调和解释上面列举的那些抽象概念的现实意义，并且使读者相信这些概念全都是既从它们自身的起源方面也从实际应用方面同生活联系着的.

不过，抽象并不是数学独有的属性，它是任何一门科学乃至全部人类思维都具有的特性. 因此，单是数学概念的抽象性还不能说尽数学的特点.

数学在它的抽象方面的特点还在于：第一，在数学的抽象中首先保留量的关系和空间形式而舍弃了其他一切. 第二，数学的抽象是经过一系列阶段而产生的；它们达到的抽象程度大大超过了自然科学中一般的抽象. 我们将以数学的基本概念：数与形为例来详细解释这两点. 最后——这也是惹人注意的——数学本身几乎完全周旋于抽象概念和它们的相互关系的圈子之中. 如果自然科学家为了证明自己的论断常常求助于实验，那末数学家证明定理只需用推理和计算.

当然，数学家们为了发现自己的定理和方法也常常利用模型，物理的类比，注意许多单个的十分具体的实例等等. 所有这些都是理论的现实来源，有助于发现理论的定理，但是每个定理最终地在数学中成立只有当它已从逻辑的推论上严格地被证明了的时候. 如果一个几何学家报告一条他所发现的新定理时，只限于在模型上把它表示出来，那么任何一个数学家都不会承认这条定理是被证明了. 对于证明一个定理的要求从中学的几何课程中就可以很好地了解到了，这种要求贯穿在全部数学中. 我们可以极精确地测量成千个等腰三角形的底角，但这并不能给我们以关于等腰三角形两底角相等的定理的数学证明. 数学要求从几何的基本概念推导出这个结果（现在在几何的严格叙述中基本概念的性质是精确地表述在公理中），并且总是这样的：证明一个定理对于数

学家来说就是要从这个定理中引用的那些概念所固有的原始性质出发，用推理的方法导出这个定理．这样看来，不仅数学的概念是抽象的、思辨的，而且数学的方法也是抽象的、思辨的．

数学结论本身的特点具有很大的逻辑严格性．数学推理的进行具有这样的精密性，这种推理对于每个只要懂得它的人来说，都是无可争辩和确定无疑的．数学证明的这种精密性和确定性人们从中等学校的课程中就已很好地懂得了．数学真理本身也是完全不容争辩的．难怪人们常说："像二乘二等于四那样的证明"．这里，数学关系式 $2\times2=4$ 正是取作不可反驳、无可争辩的范例．

但是数学的严格性不是绝对的，它在发展着；数学的原则不是一劳永逸地僵立不动了，而是变化着的并且也可能成为甚至已经成为科学争论的对象．

归根到底，数学的生命力的源泉在于它的概念和结论尽管极为抽象，但却如我们所坚信的那样，它们是从现实中来的，并且在其他科学中，在技术中，在全部生活实践中都有广泛的应用；这一点，对于了解数学是最主要的．

数学应用得非常广泛也是它的特点之一．

第一，我们经常地、几乎每时每刻地在生产中、在日常生活中、在社会生活中运用着最普通的数学概念和结论，甚至并不意识到这一点．例如，我们计算日子或开支时就应用了算术，而计算住宅的面积时就运用了几何学的结论．当然，这些结论都是十分简单的，不过，记起这一点是有益的：在古代某个时候，这些结论曾经是当时正在萌芽中的数学的一些很高的成就．

第二，如果没有数学，全部现代技术都是不可能的．离开或多或少复杂的计算，也许任何一点技术的改进都不能有；在新的技术部门的发展上数学起着十分重要的作用．

最后，几乎所有科学部门都多多少少很实质地利用着数学．"精确科学"——力学、天文学、物理学、以及在很大的程度上的化学——通常都是以一些公式来表述自己的定律（这是每个从中学

毕业的人都早已懂得的)，都在发展自己的理论时广泛地运用了数学工具．没有数学，这些科学的进步简直是不可能的．因此，力学、天文学和物理学对数学的需要恰好也总是在数学的发展上起了直接的、决定性的作用．

在其他科学中数学起着较小的作用．但是就在这些领域中，它也有重要的应用．当然，在研究像生物现象和社会现象那样复杂的现象时，数学方法本质上不能起像在物理学中所能起的那样的作用．数学的应用总是只有与具体现象的深刻理论相结合才有意义，在这些现象的研究中尤其如此．记住这一点是很重要的，这样才不致迷惑于毫无实在内容的公式游戏．但是无论如何，数学几乎在所有科学中，从力学到政治经济学，都有着这样那样的应用．

我们来回忆几个在精确科学和技术中特别出色的数学应用的例子．

太阳系最远的行星之一的海王星是在1846年在数学计算的基础上被发现的．天文学家阿达姆斯和勒未累分析了天王星的运动的不规律性，得出结论说：这种不规律性是由其他行星的引力而发生的．勒未累根据力学法则和引力法则计算出这颗行星应该位于何处，他把这结果告诉了观察员，而观察员果然从望远镜中在勒未累所指出的位置上看到了这颗星．这个发现不仅是力学和天文学特别是哥白尼体系的胜利，而且也是数学计算的胜利．

另一个同样令人信服的例子是电磁波的发现．英国物理学家马克斯威尔概括了由实验建立起来的电磁现象规律，把这些规律表述为方程的形式．他用纯粹数学的方法从这些方程推导出可能存在着电磁波并且这种电磁波应该以光速传播着．根据这一点，他提出了光的电磁理论，这理论以后被全面地发展和论证了．但是，除此以外，马克斯威尔的结论还推动了人们去寻找纯电起源的电磁波，例如，由振动放电所发射的电磁波．这样的电磁波果然被赫芝所发现．而不久之后，波波夫就找到了电磁振荡的激发、发送和接收的办法，并把这些办法带到许多应用部门，从而为全部无线

电技术奠下基础．在已成为公共财富的无线电的发现中，纯粹数学推论的结果也起了巨大的作用．

科学就是这样从观察，比如观察到由电流而引起磁针偏转，进入概括，进入现象的理论，进入规律的提出以及它们的数学表达式．新的结论从这些规律中产生，而最后，理论又体现在实践中，实践也给予理论以向前发展的新的强有力的动力．

特别值得注意的是，没有从自然科学或技术方面来的直接推动，而仅从数学本身内部产生的最抽象的数学体系，甚至也有极有价值的应用．例如，虚数在代数中出现了，在很长一段时间中它们的实在意义却没有被理解，这一情况可以从它们的名称中看出．但是以后，就在上世纪初对它们给予了几何的解释（见第四章§2），从而虚数在数学中完全站住了，并且建立了复变数（就是$x+y\sqrt{-1}$形式的变数）函数的广泛理论．这种所谓"虚"变数的"虚"函数的理论完全不是虚假的，而是解决许多技术问题的很现实的工具．比如，茹可夫斯基关于机翼上升力的基本定理正好就是以这个理论作为工具来证明的．又如，就是这个理论在解决堤坝渗水问题时也显示了它的用处，至于这个问题的意义在巨大的水电站建设时代是很显然的．

非欧几里得几何是另一个同样光辉的例子[1]．它是从欧几里得时代起的几千年来人们想要证明平行公理的企图中，也就是说，从一个只有纯粹数学趣味的问题中产生的．罗巴切夫斯基创立了这门新的几何学，他自己谨慎地称之为"想象的"，因为还不能指出它的现实意义，虽然他相信是会找到这种现实意义的．他的几何学的许多结论对大多数人来说非但不是"想象的"，而且简直是不可想象和荒诞的．可是无论如何罗巴切夫斯基的思想为几何学的新发展以及各种不同的非欧几里得空间的理论的建立打下了基础；后来这些思想成为广义相对论的基础之一，并且四维空间非欧几里得几何的一种形式成了广义相对论的数学工具．于是，至少

1）我们只在这里指出这个例子，不进行解释，读者可以在第三卷第十七章找到这个例子的解释．

看来是不可理解的抽象数学体系成了一个最重要的物理理论发展的有力工具．同样地，在原子现象的近代理论中，在所谓量子力学中，实际上都运用着许多高度抽象的数学概念和理论，比如，无限维空间的概念等等．

不必陷于例子的列举；我们已经足够地强调了数学在日常生活实践中，在技术中，在科学中都有最广泛的应用，并且只从数学本身内部生长起来的理论在精确科学和许多技术问题中也有其应用．除了数学的抽象性、严格性和它的结论的确定性以外，数学的另一个特征便是如此．

2．注意了所有这些数学的特点，我们当然还没有阐明数学的本质，毋宁说只是指出了数学的外表特征．问题在于要解释这些特点．为此至少应该回答下列问题：

抽象的数学概念反映什么东西？换句话说，数学的现实对象是怎样的？

为什么抽象的数学结论如此令人确信无疑，而原始的概念又如此显然？换句话说，数学方法的基础是什么？

为什么数学尽管如此抽象，却有最广泛的应用，而不是空洞的抽象把戏？换句话说，数学的意义从何而来？

最后，什么样的力量推动数学发展，使它把抽象性和应用的广泛统一起来？换句话说，数学发展过程的内容是什么？

回答了这些问题，我们就可以得到关于数学的对象，关于它的方法的根据，关于它的意义和发展的一般概念，也就是说抓住了它的本质．

唯心主义者和形而上学者们不但在解决这些问题方面陷于混乱，而且简直是把数学翻转过来完全加以歪曲．例如，看到数学结论的高度抽象性和明确性，唯心主义者们就想象说，数学是从纯粹思维中产生的．

事实上数学没有给唯心主义和形而上学以任何根据；恰好相反，客观地考察一下全部数学的关系和发展，它正可以给辩证唯物主义提供又一个光辉明证，并且每一步都反驳了唯心主义和形而

上学. 我们只要试图从最一般的特点上来回答前面所提出的关于数学本质的问题, 就会相信这一点的. 我们也相信对于这些问题的答案已经包含在由马克思主义经典作家所建立的关于数学以及关于科学和认识一般的本质的原理中了 为了预先解释这些问题, 考察一下算术和初等几何的基础就够了. 我们就要开始讨论它们. 进一步深入到数学中去, 当然可以加深和发展已经得到的结论, 但无论如何不会改变这些结论.

§2. 算　术

1. 关于数的概念(我们现在说的只是正整数)——这个我们如此习惯的概念, 形成却很慢. 我们根据不久以前还处于原始公社制度各个不同阶段的那些民族进行计算的情况就可以判定这一点. 有些民族甚至还没有大于二或三的那些数的名称, 有些民族虽然还可以往下多数几个数, 但无论如何还是很快就完结了, 他们把较大的数简单地称作"许多"或"无数地". 由此可见明显地划分开来了的各个数是人们逐渐地积累起来的.

起初, 人们没有数的概念, 即使他们能够用自己的方式判断出在实践中遇到的这一物体集合或那一物体集合的大小. 应该认为, 数已被他们直接了解为物体集合的不可缺的性质, 但是他们还没有明显地把数分离出来. 我们已经这样地习惯于计算, 以至不见得能表明它, 却可以理解它[1].

在更高的发展阶段上, 数已被指明为物体集合的性质, 但是还没有把它当作"抽象的数", 当作同具体物体无关的一般的数而与物体集合分离开来. 这可以从有些民族给予数的名称中看出来, 例如: "手"就是五, "整个人"就是二十等等. 这里五不是抽象地被理

1) 因为, 任一物体集合, 比如有一群绵羊或者一堆木柴, 可以从其全部具体性及复杂性中直接地被感觉出来, 而从其中分离出个别性质和关系则是经过一定分析的结果. 原始的思维还不能作出这样的分析, 而只是整体地看待客体. 正如一个不懂音乐的人欣赏音乐作品, 分不出其中细致的旋律、曲调等等, 但一个音乐家甚至能够很容易地分析复杂的交响乐.

解，而是简单地理解为"就是象手上的指头那样多"，而二十被理解为"就是象一个人身上所有的手指和脚指那样多"等等．这和有些民族还没有比如"黑色的"、"坚硬的"、"圆的"等等概念的情形完全一样．为了说明一个物体是黑色的，他们把它，例如与老鸦比较；而为了说明有五个东西，他们就把这些东西直接地与手比较．常常是这样，即数有不同的名称，用于不同种类的物体，一些是用来计算人的，另一些是用来计算船只的等等，共达数十种不同的数．这里不是抽象的数，这些"有名数"好象是分别属于一定种类的物体的．有一些民族根本没有数的独立名称，例如没有"三"字，但是他们能够说"三个人"，在"三处地方"等等．

正如我们常常说这个或那个物体是黑色的，但是很少说"黑"本身，因为这个概念比较抽象[1]．

物体的数目是物体的某个集合的性质，数本身，换句话说即"抽象的数"，是一种脱离了具体物体集合的象"黑"、"坚硬性"等那样本身已经可以设想的性质．就象黑是具有煤的颜色的各种物体的公有性质一样，数"五"是所有包含象手上的指头那样多个物体的集合的公有性质．于是等数性由简单的比较建立起来了，取出集合的一个物体，我们就弯一个手指头，就这样地用手指头一个个数出它们．完全不利用数，就用把两个集合的物体逐一比较的办法一般地即可以断定它们的物体数是否相等．例如，客人入座了，没有作任何计算，可是如果女主人少摆了一付餐具，她却能很容易地查出来，因为一个客人还没有餐具．

这样一来，可以提出数的下述定义：**每一个单个的数，象"二"、"五"等等，是物体集合的一种性质．这种性质对于所有那些物体集合之间可以将其物体逐一对比的集合来说是共同的，对于那些不能将其物体逐一对比的集合来说是不同的．**

1）关于物体性质的概念，比如物体的颜色或数目，其形成过程，可以分成三个阶段，当然这些阶段的划分也不能太严格．在第一阶段，性质是由物体的直接比较确定的：就是象老鸦这个样子；就是象手上的指头那样多．在第二阶段，出现了形容词：黑色的石头，同样地出现了数词：五株树等等．在第三阶段，性质脱离了物体，可以变成为性质"本身"，象"黑"，象抽象的数"五"等等．

为了要发现这种共同性质，并且把它明显地分出来，也就是为了建立这个数或那个数的概念并给于它名称象“六”、“十”等等，就必须在不少物体集合之间进行比较、人们世世代代地进行计算，千百万次地重复这同一种运算，于是在实践中发现了数及数之间的关系.

2. 对于数进行计算、运算，也正是对具体物体作实在计算的反映. 这也可以从数的名称中明显地看出. 例如，有些印第安人把数“二十六”说成“我们在两个十上面加上六”. 显然，这里反映出计算物体的具体方法. 尤其明显的是数的加法相当于把两个或多个物体集合堆放在一起成为一个总合. 同样地容易看出减法、乘法和除法的具体意义(特别是乘法，可以看出它的产生无非是由于把两个或三个或更多的相同的集合加起来).

在计算过程中，人们不仅发现和掌握了单个的数之间的关系，比如，二加三等于五，并且还逐渐地建立起一般规律. 在实践中发现：和数与几个被加数的顺序无关，也就是对一定物体计算的结果与这计算按怎样的顺序进行无关(后面这种情形具体表现在“序”数与“量”数相一致：第一、第二等等与一、二等等相一致). 因此数不是一个个无关的而是处在相互关联之中.

一个数在其名称及写法上甚至可通过其他数表示出来. 例如，“二十”表示“二个十”，按照法文，80——“四-二十”(quatrevingt)，90——“四-二十-十”，又如，罗马数字 VIII，IX 表示 $8=5+3$，$9=10-1$.

总之，不单是产生了一些单个的数，而且产生了具有一定关系和规律的数的系统.

算术的对象正是具有一定关系和规律的数的系统[1]. 单个的抽象数本身不具有那种包含很多内容的性质，关于它，一般地没有多少可说的. 如果我们问，比如，数 6 的性质，那末可以指出 $6=5+1$，$6=3\cdot 2$ 以及 6 是 30 的因子等等. 但是这里数 6 处处与

1) “算术”(Арифметика) 这个字起源于希腊字“计算的艺术”(“Арифмос”——数，“Техне”——艺术).

其他数关联着[1]，因此，这个数的性质正是在它同其他数的关系之中．尤其明显的是，任一种算术运算都确定数之间的一种联系，或者换个说法，确定数之间的一种关系．

因此，算术研究的是数之间的关系．但是数之间的关系是物体集合之间的现实的量的关系的抽象形态，所以我们可以说：算术是关于现实的量的关系的科学，但是这种关系是抽象的，也就是说是在纯粹形式上加以研究的．

算术，正如我们所看到的，不是象唯心主义者们企图捏造的那样从纯粹思维中产生出来的，而是反映现实物体的特定的性质；它是由于许多世代的长期实际经验而产生的．

3．社会实践愈是宽广和复杂，它就提出愈广泛的任务．不但需要指出物体的量并以关于它们的数的思想来表示，这就已经要求形成数的概念和数的名称，而且需要学习计算所有较大的集合(比如牲畜群、交换物、到指定期限前的天数等等)，要把计算的结果确定下来和告诉别人，这就正是要求数的名称以及数字符号更加完善．

看来从文字产生之初就开始引进的数字符号在算术的发展上起了巨大的作用．并且这是引进一般数学符号和公式的第一步．下一步，即引进算术运算的符号和未知数的字母符号 (x)，则是很晚才完成的．

数的概念，象任何抽象概念一样，没有一种直接的模型，不能把它表示出来，只能加以思索．但是思想是在语言中形成的，所以没有名称也就没有概念．符号也就是名称，只不过是无声的，而且是书写出来的，它把思维在能看见的形式上再现出来．例如，如果我说"七"，你想象什么呢？大概不是七个什么样的物体，而首先是数字"7"，它成为抽象数"七"的物质的外壳．又如，18273 这样一

1) 也可以从最一般的说法中理解这一点．任一种脱离了具体基础的抽象——例如数就舍弃了具体的物体集合——"本身"是没有意义的，它只存在于与其他概念的关系中．这些关系已经包含在某种说明中，也包含在很不完全的定义中．抛开这些关系，它就失去内容和意义，也就是说，根本不存在．抽象数的概念的内容包含在数的系统的规律和关系中．

个数，显然，说出它比写出它更难，并且已经不可能完全准确地想象出物体集合的模型．所以，符号可以帮助建立那样一些不能从简单的观察和直接计算中发现的数的概念，虽然不是一下子就可以做到．在这方面有很实际的必要性：由于国家的出现，就需要征收捐税，征集军队并装备军队等等，就要求对很大的数字作运算．

因此，第一，数字符号的作用就在于它们给出了抽象数概念的简单的具体化身[1]．数学符号的作用一般也就是这样：它们给出了抽象数学概念的具体化身．例如，“+”表示相加，x——未知数，a——某一个定数等等．第二，数字符号给出了非常简单地实现各种数字运算的可能性．谁都知道，“笔算”比“心算”容易．数学符号和公式一般也有这样的意义：他们使得用计算即用一种几乎是机械的动作来代替一部分推理成为可能，而且，如果计算过程被书写下来，它已有一定的可靠性．这里一切都看得见，一切都可以检验，一切都由精确的规则所确定．为了举例，可以回想起“进点”[2]的加法或任一种代数变换，比如“移项到等式另一边并变号”．

从上面所讲的可以看出：如果没有合适的数字符号就不能将算术推向前进．尤其是如果没有专门的符号和公式简直就不可能有现代数学．

自然，人们远不是很快就能够建立起现代的、这样方便的数字书写法．从古时候起各民族随着文化的启蒙出现了各种数字符号，但是这些符号不仅在图样方面而且在许多原则方面都很少和我们现在的符号相象；例如，并不是处处都用十进位系统（比如，在古巴比仑十进位系统和六十进位系统是混用的）．在附表上提供了一

1）值得注意的是，如我们所看到的，关于数的概念是这样困难地经过很长时间才建立起来，而现在这概念连小孩都很容易掌握了．为什么呢？首先，当然是因为小孩听见和看见大人怎样经常不断地运用数字，于是他们也跟着学会了．其次是因为——这一点正是我们要注意的——小孩已有现成的字和数字符号，他们先学会了数的外形，然后才掌握它的意思．

2）“进点”是在作加法时为了不忘进位而常用的一个方法．例如 $\begin{array}{r} 36 \\ +\ 28 \\ \hline \dot{6}4 \end{array}$ 在6与8的下面写上4，再在它前面注一小点表示要进一位．——译者注

	斯拉夫数字		中国数字			希腊数字	阿拉伯数字	格鲁吉亚数字	埃及数字		罗马数字	马雅部族数字
	基利尔文	古斯拉夫文	古体	商业用	科学上用				象形文字	僧侣用[1]		
0				〇	〇							𝋠
1	а҃	Ⰰ	一	〡	𝍠	α̅	ا	ა	𓏺	[illegible]	I	𝋡
2	в҃	Ⰱ	二	〢	𝍡	β̅	ب	ბ	𓏻	[illegible]	II	𝋢
3	г҃	Ⰲ	三	〣	𝍢	γ̅	ج	გ	𓏼	[illegible]	III	𝋣
4	д҃	Ⰳ	四	〤	𝍣	δ̅	د	დ	𓏽	[illegible]	IV	𝋤
5	є҃	Ⰴ	五	〥	𝍤	ε̅	ه	ე	𓏾	[illegible]	V	𝋥
6	ѕ҃	Ⰵ	六	〦	𝍥	ϛ̅	و	ვ	𓏿	[illegible]	VI	𝋦
7	з҃	Ⰶ	七	〧	𝍦	ζ̅	ز	ზ	𓐀	[illegible]	VII	𝋧
8	и҃	Ⰷ	八	〨	𝍧	η̅	ح	ჱ	𓐁	[illegible]	VIII	𝋨
9	ѳ҃	Ⰸ	九	〩	𝍨	θ̅	ط	თ	𓐂	[illegible]	IX	𝋩
10	і҃	Ⰹ	十	十	𝍠〇	ι̅	ي	ი	𓎆	[illegible]	X	𝋪
20	к҃	Ⰻ	二十	〢十	𝍡〇	κ̅	ك	კ	𓎇	[illegible]	XX	
30	л҃	Ⰼ	三十	〣十	𝍢〇	λ̅	ل	ლ	𓎈	[illegible]	XXX	
100	р҃	Ⱃ	百	百	𝍠〇〇	ρ̅	ق	რ	𓍢	[illegible]	C	
1000	҂а	Ⱚ	千	千	𝍠〇〇〇	͵α	غ	ჩ	𓆼	[illegible]	M	

各民族的数字符号

选自巴什马科娃和尤士吉维奇的文章"计算系统的产生"(《初等数学百科全书》第一卷,1951,莫斯科版).

1)僧侣用文字是古代埃及象形文字简化而来的.——译者注

些不同民族的数字符号作为例子．其中，我们看到，古希腊人，以后还有俄罗斯人用的是字母符号．我们现代的“阿拉伯”数字和数字书写的一般方法是起源于印度的，十世纪时阿拉伯人把它们从印度传入欧洲，在欧洲经历好几个世纪最终地固定了下来．

我们的符号系统的第一个特点在于它是十进位系统．但是这一特点不是十分重要的，因为也可能很成功的采用，比如说十二进位系统，这时只须引进十和十一的特别符号就行了．

我们的符号系统的主要特点在于它是“位置的”系统，就是说在这种符号系统中，同一个数字由于它所在位置不同而有不同的值．例如，在372这个符号中，数字3表示三百，而7表示七十．这种书写方法不但简短，而且大大地便于计算．罗马符号就没有这样方便：同一个数372的罗马写法是：CCCLXXII，而乘上一个较大的数，按罗马写法就非常不方便．

数字的位置书写法要求用某种方法表明某一位是缺位，因为如果不表明它，我们将会把比如，三百零一和三十一混淆起来．在缺位处放上一个零；我们就可以区分301和31．在后期的巴比仑楔形文字中已经出现了零的萌芽形式．系统地引进零则是印度人的成就[1)]；这个成就使得印度人有可能贯彻我们现在所用的数字书写的位置系统．

但是这种情形很少：把零引进数的系统中，它也成为一个数．零本身就是什么也没有——在梵文中（古印度文）它被称为：“空的”（çūnga）．但是在和其他数的关系中零获得了内容，获得了一定的性质，例如任何一个数加上零仍等于这个数，但乘上零就等于零了．

4．现在回到古代算术．流传到现在的最古老的巴比仑和埃及数学书本要追溯到纪元前一千多年以前．在这些书本和有一些较迟的书本里包含了各种各样的算术问题并带有解答，甚至于还有象解某些二次甚至三次的方程或者级数这样一些现在属于代数

1）记载有零的最初的印度人手稿是九世纪末的，其中写有270这个数字，跟我们用的符号完全一样．但是，零的引进在印度大概还要早一些——在六世纪的时候．

范围的问题（当然，所有这些都是出现在具体问题和数字的例子中）. 从巴比仑流传到现在的还有数字的平方表、立方表和倒数表. 有人推测，那时已经积累起一些与实际问题没有直接关系的数学兴趣.

不管怎样，算术在古巴比仑和埃及得到了很好的发展. 但是它还不是关于数的数学理论，而是各种计算规则和各种问题的解答的汇集. 这样，现在在初级中学讲授的算术，能使所有不专门从事数学的人也都了解. 这是完全合理的，不过算术在这种形式上终究还不是数学理论，因为其中没有关于数的普遍定理.

向理论算术的过渡是逐渐进行的.

我们已经讲过符号给出了运用大数的可能性，这些大数已经无法一目了然地表示物体集合的形式，也无法从一开始依序地数到它们. 如果一些未开化民族的数字到 8, 10, 100 等就中断了，而把再大些的数字统统叫作不确定的"许多"，那末在中国、巴比仑和埃及，符号已提供了有超过万甚至百万以上的数的可能性. 这里就已显示出数列无限延续下去的可能性. 但是人们不是很快就明确意识到它的，究竟是在什么时候，我们现在也弄不清楚. 阿基米德（纪元前 287—212 年）早在他卓越的著作《数砂法》中就指明了命名大量砂粒的数目的方法，这些砂多到可以填满恒星大球. 命名和书写这样一种数的可能性在当时还需要很详细的解释.

希腊人在纪元前三世纪就已经明确意识到两种重要思想：首先是，数列可以无限地延续下去，其次是，不但可以运用任何给定的数，而且还可以讨论一般的数，建立和证明关于数的普遍定理. 这是前人大量运用具体数的经验的概括. 正是从这个经验中表现出关于数一般讨论的普遍规律和方法. 于是产生了向更高程度抽象的过渡：从单个给定的（虽然也是抽象的）数过渡到一般的数，到任何可能的数.

我们用对已有的数加上一的方法就从逐一计算物体的简单过程继而得到数字形成的无限过程的观念. 数列已被认为是可以无限地延续下去的，随之在数学中出现了无限这个概念. 当然，我们

实际上不能用逐个加一的方法得到数列中任意大的数，因为纵使把一百年的时间压缩得几乎比一秒钟的四十分之一还短，谁又能数到一万亿呢？但是问题不在这里．一的逐个迭加过程，任意大的物体集合的形成过程原则上是无限的，就是说，存在着数列无限延续下去的潜在可能性．对计算的实际局限性这里不予考虑．关于数的普遍定理已经涉及到这个可以无限延续下去的数列．

关于任何数的某一种性质的普遍定理已经隐讳地包含了关于单个数的性质的无限多的判断，它在质上比某些可以对单个数进行检验的特称判断更丰富，因此，普遍定理必须要用从数列形成规律本身出发的一般推理方法来加以证明．这里揭露了数学的深刻特点：数学不但以一些给定的量的关系作为自己的对象，而且还以一般的可能的量的关系，也即以无限性作为自己的对象．

在纪元前三世纪写成的著名的欧几里得的《原本》一书中，已有许多关于正整数的普遍定理，特别是有一条关于一定存在着任意大的素数的定理[1)]．

于是算术变为数的理论．它已经舍弃了具体的局部的问题而进入抽象概念和推理的范围．它成为“纯粹”数学的一部分．更正确地说，它也是具有在第一节中讲过的全部特点的纯粹数学本身诞生的契机．诚然，应该指出，纯粹数学是同时从算术和几何中产生的．此外，在算术的一般法则中已有代数的萌芽，代数以后从算术中分离出来．关于这一点我们下面将要讲到．

现在只剩下对我们的全部结论作出总结了，因为我们虽然是很粗浅地，但毕竟是研究了从数的概念萌芽开始的理论算术的产生过程．

5．因为理论算术的产生是数学产生的一部分，所以很自然地可以期望我们关于算术的许多结论也能用来说明关于一般数学的普遍问题．我们在这里重新提出这些问题，并把它们用之于算术．

(1) 算术的抽象概念是怎样产生的？反映了现实界中的什么东西？

1) 注意，除了 1 以外那些只有被它自己或 1 来除才没有余数的正整数，叫作素数．

前面说过的关于算术的萌芽的所有的话，就回答了这个问题．算术的概念反映了物体集合的量的关系．这些概念是在分析和概括大量实际经验的基础上加以抽象化而产生的，并且它们是逐渐地产生的；最初是与具体对象相联的数，然后是抽象的数，最后才是关于一般的数、关于任何可能的数的概念．每一阶段都是以应用先前的概念而积累起来的经验作准备的(顺便提一下，数学概念形成的基本规律之一也是如此：数学概念是以应用先前的抽象概念积累起来的经验为基础，通过一系列的抽象与概括过程而产生的)．

算术概念产生的历史证明：那种认为这些概念是从"纯粹思维"，从"原始直观"，从"先验形式的直观"或者从什么别的东西产生的那种唯心主义观点是完全错误的．

(2) 为什么算术结论如此令人信服和确定不移？

历史也给我们回答了这个问题．我们看到：算术结论本身的建立是缓慢地和逐渐地；它们反映了在许多世代的长期过程中积累起来的并且因此凝固在人们意识中的经验．它们在语言中固定下来了：在数字的名称中，在符号中，在对数字的同一种运算的经常重复中，以及在它们经常的实际应用中固定下来了．因此它们获得了明显性并令人信服．逻辑推理方法本身也有同样的产生过程．并且重要的不但是可重复性本身，而且是现实关系所客观具有的并且反映在算术的基本概念和逻辑推论的法则中的稳定性和明确性．

算术的令人信服的根源就在这里．它的结论逻辑地从其基本概念中推导出来，而逻辑方法和算术概念两者都是以数千年的实践为基础，以我们周围世界的客观规律性为基础在人们意识中形成和巩固起来的．

(3) 为什么算术尽管其概念是抽象的，却能有这样广泛的应用？

答案很简单，算术的概念和结论概括了大量的经验，在抽象的形式中表现出现实界的那些经常和到处碰到的关系．计算的对象

既可以是房间里的东西，也可以是星球，是人，是原子……．算术舍弃了所有局部的和具体的东西，而抽取了某些普遍的性质．正是因为它仅仅抽取普遍的性质，所以它的结论才能运用到这样大量的情况中去．因之，正是算术的抽象性保证了广泛应用的可能性(这里重要的是，这种抽象性不是空洞的，而是从长期实际经验中提取出来的)．对于全部数学，对于任何抽象概念和理论也都是这样的．理论应用的可能性取决于其中所概括的原始材料的广泛程度．

同时，任一抽象概念，包括关于数的概念，由于它本身特有的抽象性，所以在其意义上是有局限的．第一，在应用到任一具体对象时，它只反映了对象的一个方面，因此只能给出关于对象的很不全面的观念．例如，常有这样的情况：有了一些数据但是关于事情的本质却说明得很少．第二，不能没有任何条件地到处应用抽象概念，不能把算术运用到任一具体问题，而不判断一下，在这里运用算术是否有意义．比如，我们说到加法时，只是想象地把对象联合起来，而对于这些对象本身当然什么事情也没有发生．但是如果我们把加法运用于对象的实际联合，如果我们实际上把对象置放在一起，比如把它们放成一堆或都摆在桌上，那末这里发生的不是抽象的加法，而是现实的过程．这个过程不但不限于是算术的加法，而且可能使算术加法根本不适用．例如，物体迭放在一堆可能要折坏；野兽放在一起可能会有这个把那个吃掉的情况；“加在一起的”物质可能发生化学反应：一公升水和一公升酒精混合以后由于两种液体的相互溶解得出的不是 2 公升而是 1.9 公升的混合物．

还需要其他的例子吗？这样的例子可以举出随便多少来．

总之，真理是很具体的；正是由于数学的抽象性，记住这一点就特别重要．

(4) 最后，我们提出的末了一个问题是关于数学发展的动力问题．

对于算术来说，这个问题的回答从算术产生的历史上也可以

清楚地得出．我们看到：人们在实践中掌握了计算，形成了数的概念；然后实践又要求要有表示数的符号，并提出更困难的任务．一句话，社会实践是算术发展的动力．并且实践同概括了实践经验的抽象思维处在经常的相互作用之中．在实践的基础上产生的抽象概念成为实践的重要工具，并在应用中日趋完善．舍弃次要的东西有助于揭露事物的本质，并且在抽象过程中分离出来和保存下来的普遍性质和关系——在算术中即数量的关系——对于保证问题能有一般解决来说是起决定性作用的．

此外，思维常常超出实践提出的任务的直接要求以外很远．比如，象关于百万或十亿这样一些大数的概念虽然是在计算的基础上产生的，但是这些概念却先于运用这些数字的实际需要．这种例子在科学史上是不少的；只要提起我们曾说到过的虚数就够了．对于实践和抽象思维，实践和理论之间的相互作用的全面认识来说，所有这些只是普遍原理的特殊情况．

§3．几　　何

1．几何产生的历史本质上同算术产生的历史相似．最初的一些几何概念和知识也要追溯到史前时期，也是在实践活动的进程中产生的．

人从自然界本身提取出几何的形式．月亮的圆形和镰刀形，湖的水平面，光线或整齐的树木的直，都是早在人类以前就存在的，以后也时时刻刻呈现在人们面前．当然，我们在自然界中很少看到十分直的线，尤其是三角形和正方形．很明显，人们建立起关于这些图形的观念首先是因为他们主动地领会了自然界，并且按照自己实践的要求制造出具有越来越规则的形状的物体．人们建筑自己的住所，磨光石头，圈出一小块土地，拉紧自己弓上的弦，制造陶器，改善陶器，并相应地建立起各种概念，比如制造出来的器皿是圆形的，拉紧了的弦是直的．总之，起初是把形式赋予原材料，而以后就意识到形式是从属于原材料的，也是可以脱离开原材

料独立地加以考察的. 人们意识到物体的形式，就能够改进自己的手工品，并且能够更明确地把形式概念本身分离出来. 这样，实践活动成了建立几何抽象概念的基础. 需要制造成千个具有直的边沿的东西，拉紧成千条绳子. 在地上划出大量直线，才能得出关于一般直线的清晰的概念，把直线看作所有这些特殊情形的共同的性质. 现在我们周围尽是人们制造出来的具有直的边沿的各种物体，我们自己学会了划直线，所以我们从小就能有关于直线的清晰的概念.

关于几何量的概念——长度、面积、体积——也同样是从实践活动中产生的. 人们为了自己的实际目的去测量长度，确定距离，用眼睛估计面积和体积. 在这里最简单的一般规律，最初的一些几何关系，例如长方形的面积等于它的两边的乘积等渐渐地被人们发现了. 对于庄稼人来说，为了估量播种面积和预计收成，知道这样一些几何关系是很有用的.

几何就是这样从实践活动及生活里的问题中产生出来的. 正如古希腊的学者罗德的欧第姆在纪元前四世纪所写的："几何是埃及人发现的，从测量土地中产生的. 因为尼罗河水泛滥，经常冲去界限[1]，所以这种测量对埃及人是必需的. 这门科学和其他科学一样，是从人类的需要产生的，对于这一点是没有什么可惊异的. 任何新产生的知识都是从不完善的状况过渡到完善的状况. 知识通过感性的感觉而产生，逐渐成为我们考察的对象，而最后变成理性的财产."

当然，土地测量不是激起古人建立几何学的唯一课题. 从流传至今的断简残篇中可以判断这些课题的性质和古代埃及人和巴比仑人是怎样解决这些课题的. 流传至今的一本最古的埃及人的著作是纪元前一千七百多年写出来的——这是一本由某个姓阿赫美斯的人写成的"文牍员"(皇官)手册. 在书中汇集了计算容器和仓库的容量、土地面积、土工作业的多少等等一系列课题.

---

1) 这里指的是地界. 顺便指出，几何 (геометрия) 这个字译意就是测地术 (按希腊文"гэ"——土地，"метрео"——量度).

埃及人和巴比仑人会测定最简单的面积和体积，知道圆周对直径的很精确的比率，并且甚至能够计算球的表面积，总之，他们已经有了不少几何知识．但是，可以断定，他们还没有作为一门有自己的定理和证明的理论科学的几何学．正象当时的算术一样，几何基本上是从经验中提取出来的一些规则的汇集．并且，几何一般地并没有从算术中划分出来．几何问题在计算上同时也是算术问题．

纪元前七世纪时几何从埃及传到希腊，在希腊，伟大的哲学唯物主义者法勒斯、德莫克利特和其他人又将它进一步发展了．毕达哥拉斯——唯心主义的宗教哲学学派的奠基人——的门生们也对几何作出了卓越的贡献．

几何是朝着积累新的事实和阐明它们相互间关系的方向发展的．这些关系逐渐地转变为从一些几何原理得到另一些原理的逻辑推论．用这种方法首先形成了关于几何定理及其证明的概念本身，其次阐明了那些可以从中推导出其他原理的基本原理，就是说阐明了几何的公理．

几何就是这样逐渐地转变成为数学理论．

大家知道，纪元前五世纪时几何的系统的叙述就在希腊出现了，但是它们没有流传到我们手里，显然是因为它们都被欧几里得(纪元前三世纪)的《原本》一书所排除了．在这部作品中几何已被表述为如此严密的系统，以致直到罗巴切夫斯基以前，即两千多年以来原则上已不能对它的原理添加什么新东西；而老版的吉谢列夫的著名中学教科书，也和全世界其他许多老版教科书一样，不过都是欧几里得的著作的通俗改写本而已．象欧几里得的《原本》——这一位希腊天才的完美的创造物——这样长命的书在世界上是很难找到几本的．当然，数学向前发展了，我们对几何基础的认识也更加深刻了，可是欧几里得的《原本》在许多方面仍不失为纯粹数学著作的典范．在这部著作中，欧几里得对过去的发展作了总结，把他那个时候的数学表述为一门独立的理论科学，也就是说归根到底表述为我们现在所理解的那样．

2. 几何产生的历史对于那些从算术产生的历史中得出的结论提供了根据. 我们看到几何是从实践中产生的，它向数学理论的转变在很久以前就发生了.

几何从事于"几何物体"和图形的研究，研究它们的量的关系和相互位置. 但是几何物体不是什么别的东西，正是舍弃了其他性质比如密度、颜色、重量等等，而仅仅从它的空间形式[1]的观点来加以考察的现实的物体.（几何图形是更一般的概念，其中甚至舍弃了空间的延伸. 例如，曲面只有二维，线只有一维，而点根本没有维. 点是关于线的顶端，关于精确到极限位置的抽象概念，所以点已不能再划分为几部分. 顺便提一下，欧几里得已定义了所有这些概念.）

这样，几何以舍弃了所有其他性质，换句话说，即采取"纯粹形式"的现实物体的空间形式和关系作为自己的对象. 正是这种抽象程度把几何同其他也是研究物体的空间形式和关系的科学区别开来. 例如，在天文学中，研究物体的相互位置，但只是天体的相互位置，在测地学中研究地球的形式，在结晶学中研究晶体的形式等等. 在所有这些情形中，研究具体物体的形式和位置是与它们的其他性质关联着或者相互依赖着的.

抽象引起了几何的思辨方法，对于没有任何厚度的直线，对于"纯粹形式"是不能做实验的. 只有用推理的方法从一些结论导出另一些新结论. 所以，几何定理应该由推理来证明，否则这定理就不属于几何，就与"纯粹形式"无关.

几何的原始概念的明显性，推理的方法，它的结论的令人信服都象在算术中那样有同样的起源. 几何概念的性质，如同概念本身一样，是人们从周围自然界中抽象出来的. 人们许多次地划直线，然后才能够领会通过任意两点可以划一条直线这个公理；人们千百万次地把各种物体相互搬挪和迭合，然后才能够把这种情况概括为几何图形迭加的概念，并且把这个概念应用到定理的证明上（例如在两个三角形相等的著名定理中所作的那样）.

---

1) 在形式中也包括大小.

最后谈谈几何的普遍性．球的体积等于 $4\pi R^3/3$，不管研究的是球形容器，是钢球，是星球，还是水滴等等．几何能够把所有物体的共有的东西划分出来，是因为任一现实的物体都多少具有一定的形式、大小以及相对于其他物体的位置．所以不难理解，几何的应用几乎也和算术一样广泛．测量零件或看图样的工人，测定离目的地的距离的炮兵，测量土地面积的农民，估计土方体积的建筑师——他们全都利用几何的初等原理．领航员、天文学家、测地学家、工程师、物理学家都需要很精确的几何结论．

著名的结晶学家和几何学家费德洛夫的研究工作提供了用抽象几何解决自然科学中重要问题的鲜明例子．他首先提出了要找晶体的所有可能对称形式的问题，这是理论结晶学的基本问题之一．为了解决这个问题，费德洛夫舍弃了晶体的全部物理性质，而只是把它当作几何物体的规则系统（代替具体的原子系统）来加以考察．这样一来，问题就在于只要寻求那些几何物体系统所有可能的对称形式．费德洛夫把这个纯粹几何问题解决得很彻底，找出它们全部的对称形式是 230 种．同时，可能的对称形式问题的解决也是对几何的巨大贡献，为许多几何研究工作奠下了基础．

在这个例子里，如同在几何的全部历史上一样，我们看到了几何发展的主要动力．这就是——实践和抽象思维的相互作用．由观察晶体而产生的晶体对称问题是抽象地提出的，并且引起了一门新的数学理论——规则系统理论，或者是所谓费德洛夫群[1]．以后这理论本身不但在晶体的观察中得到光辉的确证，而且成为结晶学发展的普遍指南，它还推动了实验方面以及纯粹数学方面的研究工作．

§4．算术和几何

1．到现在为止，我们分别地考察了算术和几何以及它们彼此间的关系，从而数学理论间的一般关系我们都没有加以注意．然而这关系却有特别重大的意义．理论彼此间的渗透把数学推向前进，并揭示出这些理论所反映的现实界关系的丰富多采．

算术和几何不仅互相应用，而且是产生进一步的一般概念、方

1）见第三卷第十五章．

践和理论的来源．归根到底算术和几何是数学成长的两个根源．它们之间的相互作用在它们本身刚刚萌芽的时候就有了．简单的长度测量就已经是算术和几何的结合．测量物体长度的时候，我们把某种长度单位置放在物体上面，然后数一数一共置放多少次；第一步手续（置放）是几何的，第二步（计算）是算术的．又如走路时数脚步，每数一步都结合了这两种手续．

任何一种量的度量一般地都把计算同对这一种量所实行的某种专门手续结合起来．提起用量杯测量液体或者用摆的振荡次数来测量时间这两个例子就足够了．

但是在测量的时候发现：一般说来所选用的单位不能在被测的量上置放整数次，所以在测量时不能满足于简单计算有多少单位．这时必须把单位加以分划，以便利用单位的一部分来更准确地表示量，就是说不是用整数，而是用分数来表示量．分数事实上就这样产生了，正如从有关的历史材料和其他材料的分析中所指明的那样：它们是从连续量的分割和比较中，即从度量中产生的．人们度量的最初一些量就是几何量：长度、播种面积、液体和散粒体的体积．我们已经在分数的产生过程中看到算术和几何的相互作用．这种相互作用引起了分数这个重要新概念的出现，引起了数的概念从整数到分数的推广（或者象数学家所说的，推广到用整数之比来表示的有理数）．分数不是也不能从整数的分割中产生，因为整的对象要用整数来计算．三个人，三支箭等等——这是完全有意义的，但是两个三分之一的人或三分之一支箭都是毫无意义的；甚至用三个三分之一支箭也不能射中一只鹿，因为要达到这目的必须用整的箭．

2．在与算术和几何的相互作用密切有关的数的概念的发展过程中，分数的出现只是第一步．下一步是不可通约的线段的发现．两个线段叫作不可通约的，就是如果没有一个线段能够在这两个线段上面都置放整数次，换句话说，就是这两个线段的比不能用普通分数来表示，也即不是整数之比．

起初人们简直没有很多地考虑是否任一长度都可由分数表

示．如果分割和度量达到过于细小的程度时，那末这些细小的部分就简单地给略掉了，因为度量的无限精确在实际上是没有意义的．德莫克利特甚至提出了这样的概念，即几何图形是由特种原子组成的．按照德莫克利特的说法，线段是原子的行列，因此线段之比就是它们之中的原子数目的比，也就是说可以明显地用分数来表示．这个概念，在我们看来，可能是非常奇怪的，但是对测定面积和体积却很有用处．面积的计算就是由原子所组成的行列的总和，体积就是原子层的总和．德莫克利特用这种方法求出了，比如说，圆锥体的体积．懂得积分概念的读者容易发觉：在这个方法中已包含了用积分法来确定面积和体积的征兆．（此外，当我们想象德莫克利特那个时代时，应该努力摆脱那些已被数学的发展所证实的在现在已成为习惯的概念．在德莫克利特的时代，几何图形脱离现实还没有达到我们现在这样的程度．因此，既然德莫克利特设想物体由原子所组成，自然应该认为几何图形也是由原子所组成的．）

但是，线段由原子组成这个概念与毕达哥拉斯定理（即勾股定理）发生了矛盾，因为从这个定理推出不可通约线段的存在；例如，正方形的对角线与它的边是不可通约的，就是说它们的比不能用整数的比来表示．

我们来证明，正方形的边和对角线的确不可通约．设 a——正方形的边，b——正方形的对角线，那末按照毕达哥拉斯定理 $b^2=a^2+a^2=2a^2$，所以，$\left(\frac{b}{a}\right)^2=2$．

但是却没有这样一个分数，它的平方等于2．因为，假定不然，令 p 和 q 表示两个整数并且有 $\left(\frac{p}{q}\right)^2=2$，我们当然可以假设 p 和 q 已经没有公共因子，因为有的话这分数可以把它约掉．但是如果 $\left(\frac{p}{q}\right)^2=2$，那末 $p^2=2q^2$，所以 p^2 可以被2除尽．在这种情况下 p^2 就可以被4除尽，因为它是整数的平方．因此 $p^2=4q_1$，即 $2q^2=4q_1$ 且 $q^2=2q_1$．由此可见，q 也可以被2除尽，这个结果当

然与 p 和 q 没有公共因子的假设矛盾. 这个矛盾证明: 关系式 $\frac{b}{a}$ 不能用有理数来表示. 正方形的对角线和边是不可通约的.

这个发现对希腊学者们造成了极大的印象. 现在, 我们已习惯于无理数, 不费力地运用平方根或其他的根, 不可通约线段的存在也毫不使我们困窘. 但是在纪元前五世纪, 这种线段的发现对于希腊学者却完全是另外一种样子. 要知道他们没有无理数的概念, 类似 $\sqrt{2}$ 这样的符号他们也没有写过, 因此对他们说来, 所得到的结果就意味着正方形的对角线和边的比根本不能用任何数表示出来.

不可通约线段的存在, 对希腊人揭露了一个包含在连续性中的秘密——奠基于连续性和运动的一种辩证矛盾的表现. 许多知名的希腊哲学家从事于对这个矛盾的探讨, 他们中埃利亚的齐诺尤以自己的悖论著称.

希腊人建立了线段 (及一般的量) 之间的比的理论[1], 估计到有不可通约线段存在, 在欧几里得的《原本》一书中叙述了它, 在现在的中学几何课程中也以简化的形式讲述它. 但是希腊人并不能提升到这样的思想, 即理解到一个线段对另一个被取作单位的线段、即随便一个线段长度之比也可以看作是数并这样来把数的概念本身加以推广, 因为无理数的概念在他们那里还没有产生[2]. 这个工作是在较晚时期由东方的数学家们完成的; 而关于实数的不直接依据于几何的严格的数学的一般定义只是不久以前 (在上世

1) 这个理论是纪元前四世纪的希腊学者欧桑克斯创造的.

2) 由于关于量的度量的学说不属于算术的范围, 而进入了几何的范围, 所以在希腊人那里几何就是全部数学. 比如解二次方程这样一些问题, 我们现在是用代数来处理的, 而他们却是从几何上提出和加以解决的. 欧几里得的《原本》一书中包含不少这样的课题, 并且看来这部书当时给人们提出的不仅是我们现在所理解的几何的概要, 而且也是一般数学的概要. 几何的这种统治地位一直继续下来, 直到笛卡儿就反过来使几何从属于代数了. 而几何的统治地位的痕迹还保留在象二次和三次的名称"平方"和"立方"中: "a 立方"——这就是边为 a 的立方体.

纪七十年代)才给出的[1]. 从线段之比的理论建立的时候起经历了这么长的时间就说明:抽象概念的产生和精确地形成是多么困难.

3. 为了说明实数概念的特征, 牛顿在他的《数学原理》中写道:"我们与其把数理解为单位的集合, 不如把数理解为某个量对另一个被取作单位的量的抽象的比." 这个数(即比)可以是整数、是有理数,或者当给定的量与单位不可通约时,是无理数.

实数按其原来的意义无非就是一个量对另一个被取作单位的量的比;在特殊情形下——是线段的比, 也可能是面积、重量等等的比.

从而实数是脱离了具体性质来加以考察的一般量的比.

正如抽象的整数不是单个的,而是处在相互关系中、在整数的系统中才成为数学的对象一样, 抽象的实数也只是处在相互关系中,即在实数的系统中才有自己的内容并且成为数学的对象.

在实数理论中,也和在算术中一样,首先要定义数的运算:加、减、乘、除, 以及用"大于"和"小于"来表示的数之间的关系. 这些运算和关系反映出各种量的现实关联,例如,加法反映了线段的相加. 中世纪的东方数学家为抽象的实数的运算奠定了基础. 以后逐渐地提出了实数系统的最重要的性质——它的连续性. 实数系统——这是连续变化着的量的一切可能值的抽象模型.

这样,正如整数的算术一样,实数的算术以连续量的现实的量的关系作为自己的对象, 它是在普遍的形式上完全脱离任何具体性来研究这些连续量的. 正因为在实数的概念中提出了为所有连续量所共有的东西,所以它有这样广泛的应用:各种不同的量,比如长度、重量、电流强度、能等等的值都可以由数来表示, 而各量之间的依存和联系关系, 可以用它们的数值之间的依存关系来刻划.

---

1) 这里说的不是描述性的定义,而是在研究实数性质时作为证明的直接基础的定义. 当然,这样的定义在比较晚的时候才产生,这时数学的发展,主要是无穷小量的分析,要求有相应的实数"变量 α"的定义. 这个定义在上世纪七十年代由德国数学家魏尔斯特拉斯、戴德金和康托尔分别在不同形式中给出.

为了能够使关于实数的一般概念成为数学理论的基础，就必须给出它们的数学形式的定义．这可以用各种方法作出来，但是，大概从量的度量过程本身引出来是最自然不过的，量的度量过程也正好是关于数的概念的概括的实际起源．我们将要讨论线段的长度，当然，读者容易看出，如果讨论任何其他的可以无限分割的量也是一样的．

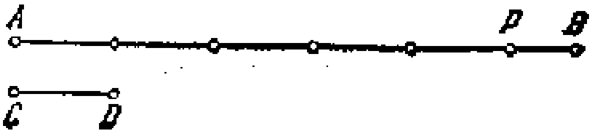


图 1

假定我们用取为单位的线段 CD 来度量线段 AB（图 1）．我们将 CD 放在 AB 上，比如从 A 点放起，它还能接着放．设它在 AB 上共放了 n_0 次．而线段 AB 还有剩余 PB，那末可以将 CD 分成十分并以这十分之一段去量剩余的线段 PB．设这十分之一段在 PB 上共放了 n_1 次．如果在这之后还有剩余，那末可以将这十分之一段再分成十分，即把 CD 分成一百分，再重复进行同样的手续……度量的过程或者结束，或者继续下去．但是不论哪种情况我们都将得出结果：整线段 CD 在 AB 上放了 n_0 次，它的十分之一——n_1 次，百分之一——n_2 次，等等．一句话，我们将得到 AB 对 CD 的比并带有很大的准确性，准确到十分之一，百分之一等等．这个比本身应该是带有整数 n_0，十分之一数 n_1 等等的十进位小数．这个小数

$$\frac{AB}{CD}=n_0.n_1n_2n_3\cdots\cdots$$

可能是无穷的，则意味着度量的无限精确的可能性．

因此，线段（和一般量）的比总能以（有穷的或无穷的）十进位小数来表示．但是在十进位小数中已经没有具体量本身的痕迹．所以它正是给出了抽象的比，即实数．所以实数形式上可以由有穷的或者无穷的十进位小数来定义[1]．

为了把工作进行到底，还需要定义对小数的运算（加法等等）．这是用这样的方法完成的，即使得定义出来的对小数的运算与对量的运算相适应．这样，当线段相加时就是把它们的长度加起来，线段 $AB+BC$ 的长度等于 AB 和 BC 的长度的和．定义对实数的运算有这样一种困难，即这些数一般说来是无穷小数，而普通已知的运算规则是属于有穷小数的．因为这个缘故，对无穷小数的运算的严格定义是用以下方法作出的．例如，我们要把 a 和 b 两个数相加．就取两个具有相同准确位数的小数，比如，都准确到百万分之一，然后将它们加起来．这时得到的和 $a+b$ 也有相应的准确度（准确到百万分

1）不考虑那些带有循环节 9 的小数，按照已知的规则可以把它们和没有循环节 9 的相应小数看成同一个数，从 0.13999…＝0.14000… 这个例中可以明显地看出这个规则。

之二，因为 a 和 b 的误差可能要加起来)．这样一来，我们就可以定义具有任意准确度的两个数的和，并且在这样的意义下它们的和是完全确定的，虽然在计算的每一步上只知道它带有某种准确度．但是，这符合于事情的本质，因为每一个量 a 和 b 都被度量到某种准确程度，而以无穷小数来表示的精确值是作为量之值的可能无限精确化的结果而得出的．

"大于"，"小于"等关系可以随后通过加法来定义：如果有这样一个数 c，使得 $a=b+c$ (这里说的是正数)，那末 $a>b$．

实数序列的连续性表现在这里：如果数 $a_1, a_2, \cdots$ 是递增的，而 $b_1, b_2 \cdots$ 是递减的，并且 b_i 总是比 a_i 大，那末无论如何在这两串数之间总有某个数 c 存在．如果直线上的点按一定的规则与数相对应，那末从直线上所画的可以很清楚地看出这一点 (图 2)．显然，数 c 的存在和对应于它的点正好表示数列中不会有间断，这就是数列的连续性．

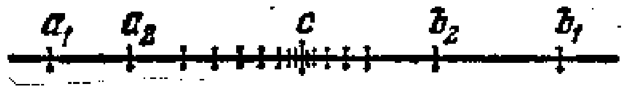


图 2

4. 在算术和几何相互作用的例子中已经可以看出，数学的发展是在交织着许多对立面的斗争过程中进行的，这些对立面是：具体与抽象，特殊与一般，形式与内容，有限与无限，连续与不连续等等．

例如，我们试着在实数概念的建立中探讨一下具体与抽象的对立．正象我们所看到的，实数反映可以无限精确化的度量过程，或者稍微理解得不同一点，即反映量的绝对的无限的精确值．这与下述情况相适应，即在几何中完全舍弃了具体对象的现实形式和大小的可变性和不确定性，而考察物体的理想的精确形式和大小．前面我们讨论的正是理想线段的度量问题．

但是理想的精确几何形式和量的绝对精确值都是一种抽象．任何具体的对象都没有绝对精确的形式，正象任何具体的量既不能绝对精确地度量出来，也没有绝对精确的值一样．例如，如果把尺的长度精确到超过原子大小的限度以外就没有意义了．量的精确化超过一定限度，总是要产生质的变化，使量一般地失去它原来的意义．比如，气体的压力不能精确到超过一个分子的撞击力的限度以外，当精确到电子的电荷大小的程度时，电荷就不再是连续的．因为自然界中的对象没有理想的精确形式，所以关于正方形

的对角线与边的比是 $\sqrt{2}$ 的断言，即不能从直接的度量中绝对精确地得出来，也不能对任一个具体的现实的正方形有绝对精确的意义.

正如我们所看到的，关于正方形的对角线和边的不可通约性的结论是从毕达哥拉斯定理推导出来的. 这是从经验材料的发展中得出的理论性结论；它是把逻辑应用于从经验中提取出来的原始几何前提的结果.

这样，不可通约线段的概念，尤其是实数的概念不是经验事实的简单的直接的反映，而是比经验事实更进了一步. 但是，这是可以理解的. 实数不是反映某种给定的具体的量，而是反映脱离了任何具体性的一般的量，换句话说，它是反映了现实的各种特殊量的共同性质. 这种共同性质就在于这些量的值一般地可以精确确定，并且如果舍弃具体的量，那末可能精确化的限度(这限度依赖于量的各种具体性质)就将成为不确定的并且会消失掉.

所以，脱离开量的个别性质来考察量的数学理论，必然应该考察量之值的无限精确化的可能性，因而应该得出实数的概念. 并且数学反映的只是各种不同量的共同性质，不能顾及每个具体情况的特点. 正象列宁所说的："任何个别都不能完全地列入一般之中……"[1] 数学分离出共同的性质，考察那些它所制造的、明显地划分出来的各种抽象体系，而不管这些抽象体系运用的现实界限. 其所以这样正是因为这些界限不具有一般性；它们依赖于被考察的现象的具体性质，依赖于这些现象的质的变化. 所以，在应用数学的时候，必须检验应用这个或那个理论本身的根据，比如，把物质当作连续体加以考察，并且以连续量来描绘物质的性质，只有舍弃物质的原子结构时才能允许，而这只有在一定条件下和某种限度内才是可能的.

但是尽管如此，实数是对现实的连续量和过程进行数学研究的经过检验的有力工具. 实数理论由实践，由它在物理、技术和化学中的广阔的应用得到了论证. 实践证明：实数概念正确地反映

1) 列宁，哲学笔记，363 页，人民出版社，1957 年版.

了量的共同性质．但是这种正确性不是没有界限的；不能把实数理论看作某种绝对的、能够完全脱离现实界而无限抽象地发展的东西．实数概念本身将继续发展，实际上它还远不是绝对终结了的．

5． 上述的许多种对立中的另一种对立——连续与不连续的对立，它的作用也可以在数的概念的发展的例子中加以探讨．我们已经看到，分数是从连续量的分割产生的．

关于分割的问题有一个十分值得深思的笑话．老奶奶买了三个马铃薯，要把它们平分给两个孙女儿．怎么办呢？回答是：必需作成马铃薯羹．

但是，这个笑话揭露了事情的本质．单个的对象在这样的意义下是不可分的，即分割开来的对象几乎再也不是它本来的样子了，这从"三分之一个人"或"三分之一支箭"的例子中可以看得很清楚．相反地，连续的和均匀的量或物容易分开来也容易结合起来，而不失去它的本性．马铃薯羹正好就是这种均匀物的很好的例子，这种东西虽然不是划分开了的，但是实际上却容易分成许多任意小的部分．长度、面积、体积都具有这样的性质，连续性就其概念本身来说就是这样一种实际上不是划分开了的，但却具有无限分割的可能性的东西．

这样，我们碰到了两个对立面：一方面是不可分的，单个的，就是说，不连续的对象，另一方面是完全可分的，但并不是划分成各个部分的，而是连续的物体．当然，这种对立总是统一的，因为既没有绝对不可分的对象，也没有完全连续的对象．但是对象的这两个方面不仅是现实的，而且常常在某些情况下一个方面表现为决定性的，在另一些情况下——另一方面是决定性的．

数学，既然使形式脱离它们的内容，因之也就把连续和不连续这两种形式明显地分割开来．

单位是单个物体的数学模型，而单位的总和就是不连续物体的集合的数学模型．这就是所谓排除了其他一切性质的纯粹不连续性的模型，就是纯粹形式的不连续性．连续的几何图形成为连

续性的基本的、原始的数学模型，最简单的情形就是直线.

从而，在我们面前有两个对立面：不连续性和连续性，以及它们的抽象数学模型：整数和几何延伸性．度量就是这两个对立面的统一：连续的东西用单个的单位来度量．但是只有不可分的单位是不行的；必须引进原单位的分数部分．这样就产生了分数；数的概念正是由于上述对立面的统一才得到发展.

进一步在更高的抽象阶段上，出现了不可通约线段的概念，以及随之而来的是作为量的无限精确值的抽象模型的实数概念．但是这个概念不是一下子就形成的，它的长期发展道路经历了连续与不连续这两个对立面的斗争.

前面已经讲过，德莫克利特认为几何图形是由原子组成的，从而把连续归结为不连续．但是不可通约线段的发现迫使人们放弃这个观念．此后，人们就不再认为连续量由单个的元素——原子或点所组成．当时人们不会用数来表示这种量，因为这时除了整数和分数以外还不知道有其他的数.

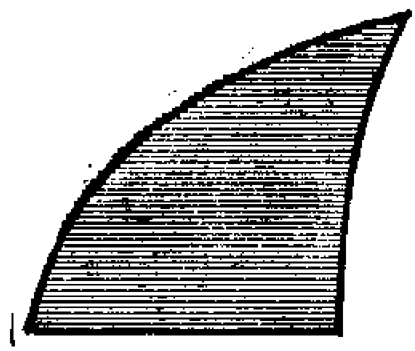


图 3

在十七世纪，当微分和积分演算的基础已经奠立的时候，数学中重新出现了连续与不连续的矛盾．这里说的是关于无穷小．在一些观念中，它们被理解为连续量的实际的“实在的”无穷小的“不可分的”部分，与德莫克利特的原子相似，不过这时认为它们的数目有无穷多．面积和体积的计算——积分——被理解为无穷多个这种无穷小的部分的总合．例如，面积被理解成“组成这块面积的那些线条的总合”（图 3）．连续再度被归结为不连续，不过已是在更复杂的形式上，更高的程度上．但是这个观点不能令人满意，而和这个观点相对立有基本上由牛顿提出的关于连续变量的概念，关于把无穷小量理解为无限减少的变量的概念．在十九世纪的前半叶，当严格的极限理论已经建立起来的时候，这个概念占了上风．现在线段不是由点或“不可分的”部分组成，而是被理解为一种延伸性，一种连续的介质，只能于

其中固定变量的单个点、单个值．数学家当时也讲“延伸性”了．在连续和不连续的统一中，连续性重又居于统治地位．

但是分析的发展要求变量理论的进一步精确化，首先是把实数看作变量的任何可能值的一般定义的进一步精确化．在上世纪的七十年代实数的理论产生了，它把线段表示为点的集合，相应地，把变量变动的区间表示为实数的集合．连续性重又由单个的不连续的点组成，而连续性的性质则表现在组成它的点的集合的构造中．这个概念使数学得到很大成就，因而取得优势．但是在这个概念中毕竟也发现了它本身的深刻困难，这困难引起了在新的阶段上重新回到纯粹连续性概念的企图．也可以用其他方法来改造把线段看作点的集合的概念．对于数、变量、函数等概念，许多新的观点在产生．理论正在继续发展，并且应该期望它的进一步发展．

6．算术与几何的相互作用不仅在实数概念的建立中起了作用．几何与算术，或者更正确地说，与代数的这种相互作用也表现在数学中负数和复数（即 $a+b\sqrt{-1}$ 形式的数）概念的确立上．负数以直线上从零起向左排列的点来表示．复数则代表平面上的点．正是这种几何表示法使虚数在数学中巩固下来，而在这以前它们一直是不可理解的．

量的概念进一步发展：例如，出现了表示有方向线段的向量及其他更一般的量（张量），在这些量中代数重新与几何结合起来．

各种数学理论的结合总是起了或者正在起着重要的、有时是决定性的作用．我们将在解析几何，微分和积分演算，复变数函数理论，最新的所谓泛函分析以及其他理论的产生的例子中进一步看到这一点．在数论本身，也即关于整数的学说中，很成功地应用了与连续性紧密有关的许多方法——无穷小量分析的方法和几何方法，产生了这个理论的称作“解析数论”和“数的几何”的庞大章节．

根据一定的观点，从数学的基础中可以看到来自几何和算术的那些概念——连续性和代数运算（作为算术运算的推广）的一般

概念的结合．但是我们不能在这里讲这些困难的理论．本节的目的只要建立起关于概念之间相互作用的最一般的概念，在算术和几何的相互作用的例子中，在数的概念的发展的例子中建立起关于在数学中对立面斗争和统一的最一般观念．

§5．初等数学时代

1．数学的发展不能归结为一些新定理的简单积累，而包含有数学的根本变化，可以说是质的变化．但是这些质的变化不是用破坏和取消原有理论的方式进行的，而是用深化和推广原有理论的方式，用以前的发展作准备而提出新的概括理论的方式进行的．

从最一般的观点来看，数学的历史可以分为四个基本的、在质上不同的阶段．当然，精确划分这些阶段是不可能的，因为每一个相继的阶段的本质特征多少都是逐渐形成的，不过这些阶段的区别和它们之间的过渡能够十分明显地表示出来．

第一个阶段（或时期）——这是数学作为一门独立的、纯粹理论的科学的萌芽时期．这个时期从最古的时代起，终止于纪元前五世纪，也许更早一些，这时在希腊最后地形成了具有理论和证明之间的逻辑关系的“纯粹”数学（特别是在纪元前五世纪出现了几何的系统的阐述，例如，希约斯的希波克拉特所著的“原理”）．这个时期是算术和几何形成的时期，我们已经足够详尽地考察过了．当时数学是作为与实践直接有关的，从经验中提取出来的许多单个法则的总合而建立起来的．这些法则还没有形成统一的具有逻辑关联的系统．数学的具有定理的逻辑证明的理论性质随着材料的积累形成得非常缓慢．算术和几何还没有分开，彼此紧密交错着．

第二个时期可以称之为初等数学即常量数学的时期；这个时期的基本的、最简单的结果成为现在中学课程的内容．这个时期延续了将近两千年，由于“高等”数学的建立而终止于十七世纪．在本节我们较详细地讲一讲这个时期．在下一节叙述第三和第四个

时期——分析的建立与发展的时代和现代数学时期.

2. 初等数学时期也可以分成基本内容不同的两个部分：几何发展的时期（到纪元二世纪）和代数优先发展的时期（从二世纪到十七世纪）. 又可以按照历史条件的不同把它分成三个时期，分别称为"希腊的"、"东方的"和"欧洲文艺复兴时代的"时期. 希腊时期正好和希腊文化普遍繁荣的时代一致；起始于纪元前七世纪，到纪元前三世纪，到最伟大的古代几何学家欧几里得、阿基米德、阿波洛尼的时代达到了自己的顶峰，而终止于纪元六世纪. 数学，特别是几何学在希腊达到惊人的繁荣. 我们知道很多希腊数学家的名字和工作成果，虽然他们的原著流传到现在的并不多. 同时值得注意的是，在纪元一世纪，兴盛的罗马并没有在数学上作出什么贡献，而那时被他们奴役的希腊，科学却更为繁荣.

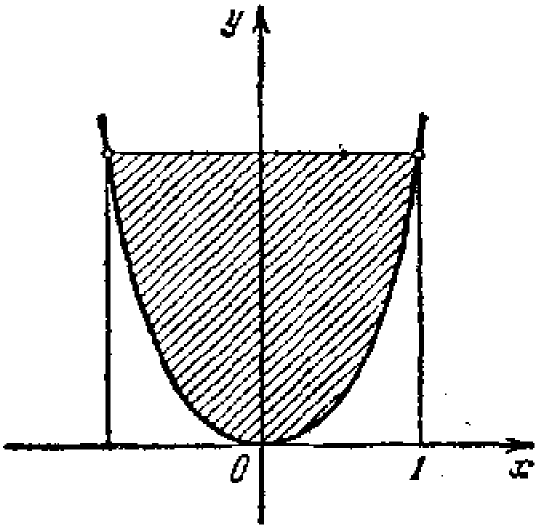


图 4

希腊人不仅在欧几里得的《原本》中所讲到的和现在中学课程内容的范围内发展了初等几何并把它导向完整的系统，还得到许多非常重要的结果. 例如，他们研究了圆锥曲线：椭圆，双曲线，抛物线；证明了某些属于所谓射影几何初步的定理；以天文学的需要为指南，建立了球面几何（在纪元一世纪）以及三角学的原理，并计算出最初的一些正弦表（吉巴尔赫——纪元前二世纪，克拉夫基·托勒密——纪元二世纪[1]）；确定了一系列复杂图形的面积和体积，比如，阿基米德确定了抛物线弓形的面积，证明它等于包住这个弓形的长方形面积的 2/3（图 4）. 希腊人甚至已经知道了这样一条定理：在所有具有给定表面积的物体中，球有最大的体积，但是这个定理的证明没有保留下来，而这个证明

1) 托勒密认为地球是宇宙的中心，星球围绕着地球运动，他是作为这样一个体系的创造者而著名的。这个体系已被哥白尼推翻了。

是这样困难，可能希腊人未必掌握了定理的完全的证明；这个证明是在十九世纪用积分演算第一次作出来的.

在算术和代数初步的领域中，希腊人也作了不少工作. 正象前面已经指出过的，他们奠立了数论的基础. 例如，他们对素数的研究(欧几里得的关于存在着无限多个素数的定理和寻找素数用的"爱拉托斯芬的筛子")以及解整数方程(丢番图，公元246—330年左右)都是属于这方面的.

我们已经说过，希腊人发现了无理量，但是是在几何上把它们当作线段加以考察的. 因此我们现在在代数上加以考察的许多问题，希腊人是在几何上考察的. 他们对解二次方程和变换无理量的表示就是这样，例如，我们现在写成 $x^2+ax=b^2$ 形式的方程，他们读作：寻找这样一个线段 x，使得以它为边作成的正方形加上以它和给定的线段 a 为边作成的长方形，就得到一个长方形，它和给定的正方形一样大. 几何的统治地位在希腊人以后还延续了很长时候. 希腊人也知道求出平方根和立方根的方法，知道算术级数和几何级数的性质.

所以，希腊人已经掌握了现代初等代数的许多材料，但是主要的材料还不够，如负数和零，脱离开任何一种几何的无理数，以及发达的字母符号系统. 不过，丢番图已经运用字母符号来表示未知数和它的方次，还有加、减号等特殊记号，因此他写出了代数方程，但是还只带着具体的数字系数.

在几何方面希腊人已经接近于"高等"数学：阿基米德——在面积和体积的计算中接近于积分演算. 阿波洛尼——在关于圆锥曲线的研究中接近于解析几何. 他实际上给出了这些曲线的方程[1)]，不过是用几何语言来表示的. 但是他们还没有任意常量和变量的一般概念，也没有那种必需的形式——代数的字母符号，这些出现于另一个时代并能使他们的研究变成进入高等数学的新理

1) 他给出了带有顶点的圆锥曲线的"方程". 例如，抛物线的"方程" $y^2=2px$，他表示成这样：以 y 为边的正方形与以 $2p$ 和 x 为边的长方形一样大(当然，他用相应的线段来代替符号 p, x, y).

论的起源．这些新理论的创立者过了千年以后在很大程度上还是以希腊学者们的遗作为指针的．笛卡儿正是从清理希腊人遗留下来的问题着手建立了解析几何基础的著作《几何学》(1637) 的．

一般规律是这样的：引起了新的深刻问题的旧理论，本身好象停步不前，而为了进一步发展要求产生新的形式和思想．这些新的形式和思想的产生又要求另外一些条件．在古代社会没有也不可能有过渡到高等数学的条件；这些条件随着自然科学在新时代的发展而出现，而这种发展在十六—十七世纪又决定于技术和工业的新的需要，因而是与资本主义的萌芽和发展有联系的．

希腊人透彻地研究了初等几何，正应该以此来解释这样一个事实，即几何的光辉发展直到现代开始以前已经穷竭，而由托勒密、丢番图等人工作中的三角和代数的发展所代替．可以认为丢番图的工作正是代数占主导地位的时期的开端．但是直到古代社会末期也还不能把科学进一步推向这新的方向．

应该指出，远在这以前好几个世纪，中国算术已达到很高的水平．中国学者们在纪元前二世纪到一世纪就叙述了三元一次联立方程组算术解的规则．同时在历史上第一次利用负系数并且表述了对负量进行运算的规则．(但是正象丢番图以后所作出的一样，他们寻找的解本身只是正的．) 在一些书中已经提出了求平方根和立方根的方法．

3. 随着希腊科学的终结，在欧洲出现了科学萧条，数学发展的中心移到了印度、中亚细亚和阿拉伯国家[1)]．在这些地方从五世纪到十五世纪的一千年中间，数学主要是由于计算的需要，特别是由于天文学的需要而得到发展；东方的大多数数学家也就是天文学家．当然，他们对于希腊的几何学几乎没有添加任何显著结果；对这门科学他们只是把希腊人的作品保留给后世．但是印度、

1) 为了表明时间起见，我们指出某些杰出的东方数学家的生活年代．印度人：阿利亚布哈大——生于 476 年左右；布拉马贡塔——约 598—660 年；巴斯加拉——七世纪；花剌子模人：阿里·花剌子模——九世纪；阿里·比鲁尼——973—1048 年；在阿塞拜疆工作过的纳西艾丁·屠西——1201—1274 年；在撒玛尔干工作过的吉雅赛金·杰姆施德——十五世纪．

阿拉伯和中亚细亚的数学家们在算术和代数领域中达到了巨大成就[1].

正如我们在§2中讲过的，印度人发明了现代计数法. 他们也引进了负数，并把正数和负数的对立与财产和债务的对立或者直线上两个方向的对立联系了起来. 最后，他们开始象运用有理量一样地运用无理量，和希腊人不同的是他们没有无理量的几何表示. 他们还有表示各种代数运算包括求根的特殊符号. 正是因为印度和中亚细亚的学者们没有对无理量和有理量的区别感到困惑，他们才能够战胜几何的"跛扈"(这正是希腊数学的特征)，从而使代数从希腊人硬加给数学的笨拙几何外壳中解放出来，打开了真正的发展道路.

伟大的诗人兼数学家奥玛尔·海扬(约1048—1122)和纳西艾丁·屠西已清楚地指明两个量(不管是可通约量还是不可通约量)的任一种比都可以称为数；于是，我们在他们那里就已经找到了数(无论有理数或无理数)的一般定义. 这个定义我们在前面§4中作为牛顿的陈述已经提到过. 如果注意到负数和无理数得到所有欧洲数学家的完全承认是很迟的事情，甚至迟到欧洲开始了数学的复兴以后，那末这些成就的伟大就表现得特别明显了. 例如，著名法国数学家维耶特(1540—1603)，代数的许多结果都归功于他，他却避开了负数. 在英国甚至到十八世纪还对负数发出抗议. 这些数被认为是荒诞的，因为它们比零还少，就是"比没有还少". 而现在负数即使表成负温度形式也成为习惯的了；我们所有的人都能从报上读到和懂得"莫斯科的温度是 $-8°$"意味着什么.

"代数"这个字本身起源于九世纪时的花剌子模数学家和天文学家穆罕默德·伊本·穆斯，阿里·花剌子模(花剌子模的穆斯的儿子穆罕默德)的作品的名称；它的关于代数的著作称作"al-Jabr w-al Mugabalab"，意思是"整理(直译为接骨)和对比. 整理——

1) 应该注意，把当时数学的发展总是与阿拉伯人联系起来是不对的，"阿拉伯的"数学这个术语多半是因为许多东方学者都采用随着阿拉伯侵略的扩展而流传开来的阿拉伯文字.

al-Jabr——就是把负项移到方程的另一边，对比——al Mugabalab——就是把方程两边的相同项消掉.

阿拉伯字"al-Jabr"翻译成拉丁文就变成了 Algebra，而 al Mugalabab 被略去了，所以出现了"代数"这个名称本身[1].

顺便提一下，这个名称的起源完全符合这门科学本身的内容. 代数的基础就是脱离了具体数字在一般形态上形式地加以考察的关于算术运算的学说. 代数的课题首先是变换表示式和解方程的形式规则. 阿里·花剌子模正是用某些一般的形式规则名称为自己的书命名，从而表示出了代数的真髓.

此后，奥玛尔·海扬把代数定义为解方程的科学. 这个定义直到上一世纪末以前都还保持着它的意义，那时在代数中除方程论外还形成了许多新方向，根本改变了代数的面貌，但是并没有改变它作为一门关于形式运算的一般学说的总精神.

中亚细亚的数学家们找到了求根和一系列方程的近似解的方法，找到了"牛顿二项式定理"的普遍公式，虽然只是用自己的语言表示它，他们有力地推进了三角学，把它建成一个系统，并且计算出非常准确的正弦表. 这些表是数学家吉雅赛金（约 1427 年）为了在乌兹别克工作过的著名天文学家乌鲁哥别克在天文学上的需要而计算出来的，吉雅赛金还发明了十进位小数，比欧洲再度发明小数早 150 年.

总之，中世纪时期，在印度和中亚细亚，现代十进位计数法（包括小数）、初等代数和三角差不多已完全地形成了. 这时中国科学的成就开始传入邻国，约在六世纪中国已经知道最简单的不定方程解法，知道几何中的近似计算以及三次方程最原始的近似解法. 从中学代数课程的材料来看，十六世纪时所缺少的大概只是对数和虚数. 此外，还缺乏字母符号系统：代数的内容胜过它的形式. 但是这种形式是必需的，因为对具体数字的抽象和一般规则的形成都要求有相应的表示方法，需要用符号来表示任何数和对于数

1）值得注意的是，表示计算方法和规则的数学术语"算法"（алгорифм）也是起源于阿里·花剌子模的名字.

的运算. 阿拉伯符号是符合于代数内容的必要形式. 正象在远古时代, 为了运用整数, 应该制定表示它们的符号一样, 现在为了运用任意数并对它们给出一般规则, 就应该制定相应的符号. 这个任务, 从希腊时代就开始解决, 而直到十七世纪才完成, 这时在笛卡儿和其他人的工作中最后形成了现代符号系统.

4. 在科学复兴时期, 欧洲人向阿拉伯学习, 并且根据阿拉伯文的翻译熟识了希腊科学. 欧几里得、托勒密、阿里·花剌子模的著作在十二世纪才第一次从阿拉伯文译成拉丁文——当时西欧的共同科学语言. 这时候通过对早先来自希腊和罗马的计数法的斗争, 从阿拉伯沿袭过来的印度计数法逐渐地在欧洲确定下来了.

只是到十六世纪, 欧洲科学终于第一次越过了先人的成就. 例如, 意大利人塔尔塔里雅和费拉里在一般形式上先解了三次方程, 然后四次方程(见第四章). (注意, 虽然这些结果没有在中学讲授, 但是按其所用方法的水平是属于初等代数的. 在高等代数中必须包括普遍的方程理论.)

在这个时期第一次开始运用虚数(当时是纯粹形式的, 没有任何现实的根据, 因为这种现实根据是在很迟以后, 直到十九世纪初才弄清楚的). 现代的代数符号也制造出来了, 其中不仅出现了表示未知数的字母符号, 也出现了表示已知数的字母符号"a", "b"等等(维耶特在 1591 年作出的). 随着代数的发展许多数学家参加了这项工作. 顺便指出, 当时在欧洲出现了十进位小数(尼德兰学者斯蒂文在 1585 年发明了并且写出了它们).

最后, 在英国纳皮尔发明了供天文计算作参考的对数, 并在 1614 年予以发表, 而布利格计算出来第一批十进位对数表, 是出现在 1624 年[1].

---

1) 指出这样一点是有趣的, 即纳皮尔所定义的对数并不象我们现在所说的: 在公式 $x=a^y$ 中数 y 是 x 以 a 为底的对数. 对数的这种定义较晚才出现. 而纳皮尔的定义同变量和无穷小量的概念联系了起来, 归结为这样的形式: x 的对数 $y=f(x)$ 是这样一个函数, 它的增长速度与 x 成反比, 即 $y'=\frac{c}{x}$ (见第二章). 所以, 本质上最原始的是定义了对数的微分方程, 虽然那时微分还没有被发明.

当时在欧洲也出现了"组合论"和"牛顿二项式定理"[1]的普遍公式；级数知道得更早．所以，初等代数的建立是完成了．同时在十七世纪初，结束了常量数学即初等数学的整个时期，现在中学里学习的就是这些初等数学，只有不多的增添：算术，初等几何，三角，初等代数的一切重要内容在当时都形成了．以后则是向高等数学——变量数学的过渡．

但是不应该设想初等数学的发展就此结束了．它仍继续发展着，例如，在初等几何中就不断出现过和将要出现许多新的结果，并且正是因为数学进一步发展，我们才能更清楚地理解初等数学的本质．但是现在变量、函数、极根等概念在数学中已具有指导的作用．从初等数学提出的许多问题现在不仅常借助于高等数学中的这些概念以及和它们有关的方法来阐明和解决，而且这些问题有时根本不可能用初等方法去解决．就是由于有了这些高等数学的概念和方法，从初等数学提出的许多问题现在成了更普遍的结果乃至理论的起源．这方面的例子有已经提到过的图形的规则系统理论或者一些数论问题，这些问题就其陈述来说是初等的，但就其解决的方法来说却完全不是初等的，关于这些例子读者可在第十章（第二卷）更详细地了解到．

§6．变量的数学

1．到十六世纪，对于运动的研究变成了自然科学的中心问题．实践的需要和各门科学本身的全部发展使自然科学转向对运动的研究，对各种变化过程和各种变化着的量之间的依赖关系的研究．

作为变化着的量的一般性质及它们之间依赖关系的反映，在数学中产生了变量和函数的概念，而数学对象的这种根本扩展就决定了向数学的新阶段——变量的数学的过渡．

1）这个公式以牛顿为名，不是因为他第一次发现了它，而是因为他把这个公式从正整数 n 推广到任意分数和无理数。

物体沿着给定的轨道运动，比如沿直线运动的规律是这样决定的，即物体走过的路程随时间而增加.

例如，伽利略（1564—1642）确定了落体所通过的路程与时间的平方成比例地增长，发现了落体的规律，这个规律表示成著名的公式：

$$s=\frac{gt^2}{2} \tag{1}$$

（这里 $g\approx 9.81$ 米/秒²）.

一般地，运动规律给出在时间 t 之内所走过的路程．这里时间 t 和路程 s 是两个变量："自变量"和"应变量"，而对应于每个时间 t 有一段确定的路程 s，这个事实就表示路程 s 是时间 t 的函数.

变量和函数这两个数学概念，无非就是具体变量（如时间、路程、速度、转动角、扫过的面积等等）和它们之间的依赖关系（如路程对时间的依赖关系等等）的抽象概括．正像实数的概念是任意量的值的抽象模型一样，"变量"是变化着的量——在加以考察的过程中必然采取不同值的量——的抽象模型．数学变量 x 无非就是"某种量"，或说得更好一些，不管哪种量，只要能取各种不同数值的量，这就是一般的变化着的量；可以把它理解为时间、路程或任何其他的量.

函数也完全一样，它是一个变量对另一个变量的依赖关系的抽象模型．y 是 x 的函数这个断言，在数学中就表示：对于每个取为 x 的值，对应着一个确定的 y 值．（这种对应本身或 y 值对 x 值的对应规律也就称为函数）例如，按照落体规律，通过的路程与降落时间由上式(1)联系起来．路程是时间的函数.

运动着的物体的能量是通过它的质量和速度按照公式

$$E=\frac{mv^2}{2} \tag{2}$$

来表示的．对于给定的物体，能量是速度 v 的函数.

在导线中有电流通过时单位时间内产生的热量按照已知的规

律以公式

$$Q=\frac{RI^2}{2} \tag{3}$$

来表示，这里 I——电流强度，而 R——导线的电阻．当电阻一定时，对于每一电流强度的值 I 对应着在单位时间内产生出来的确定的热量 Q．所以，Q 是 I 的函数．

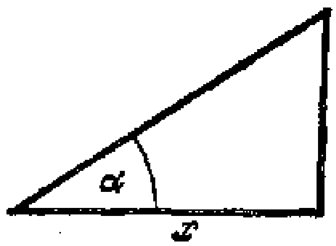


图 5

具有给定锐角 α 和直角边 x 的直角三角形的面积(图 5) 以下列公式

$$S=\frac{1}{2}x^2 \operatorname{tg}\alpha \tag{4}$$

来表示．当角度 α 确定时，面积是直角边 x 的函数．

所有的公式(1)—(4) 可以统一成一个公式：

$$y=\frac{1}{2}ax^2. \tag{5}$$

这就是从具体变量 t, s, E, Q, v 等等过渡到一般变量 x 和 y，从具体的依赖关系(1)，(2)，(3)，(4)过渡到它们的一般形式(5)．如果力学和电学研究的是与具体量联系着的具体公式(1)，(2)，(3)，那末关于函数的数学学说研究的则是与任何具体量都没有联系的一般公式(5)．

对于具体量的进一步抽象就是：不是考察像 $y=\frac{1}{2}ax^2$，$y=\sin x$，$y=\lg x$ 等等 y 对于 x 的给定依赖关系，而是考察 y 对于 x 的一般函数关系，以抽象公式

$$y=f(x)$$

来表示．这个公式的意思是：量 y 是 x 的某个函数，就是说，对于每个可以取为 x 的值，依某种方法对应着一个确定的 y 值．不但这样或那样的给定函数 ($y=\frac{1}{2}ax^2$，$y=\sin x$ 等等) 是数学的对象，而且任意的(更准确地说：多少是任意的)函数也成了数学的对象．起先是舍弃具体量，而后是舍弃具体函数的这种抽象步骤与

整数概念形成时所经历的抽象步骤相似，即起先是舍弃具体的物体集合而进入关于单个数的概念（1, 3, 12 等等），再一步抽象就进入任意整数的概念，这种概括是分析和综合，对个别关系的分析和在新概念形式中表现这些关系之一般特征的综合的深刻相互作用的结果.

数学中专门研究函数的领域叫作分析，数学分析或者有时候叫作无穷小量的分析. 这最后一个名称的来源是因为无穷小量概念是研究函数的重要工具（这概念的内容和意义将在第二章中说明）.

因为函数是一个量对于另一个量的依赖关系的抽象模型，所以可以说，分析是以变量之间的依赖关系作为自己的对象，但是不是这些或那些具体量之间的依赖关系，而是脱离了它们的内容的一般变量之间的依赖关系. 这样的抽象保证分析有广泛的应用，因为分析在一个公式中，一个定理中包括了无穷多种可能的具体情况. 简单的公式(1)—(5)就是例证. 在这里可以看出分析同算术、代数是完全类似的. 它们都是从一定的实际问题产生，都是在一般的、抽象的形式中反映出现实界的现实的量的关系.

2. 所以，从十七世纪开始的数学的新时期——变量数学时期可以定义为分析出现和发展的时期（这是以前例举过的数学发展的几个重要阶段的第三个）. 但是，大家知道，任何一种理论都不能只是由于一些新概念的形成而产生，同样地分析也不能只是出之于变量和函数的一些概念之中. 为了建立理论，尤其是建立像数学分析那样的整个科学部门，必须要使新概念动作起来，要运用它们发现新的相互关系，要使它们能够解决新的问题.

不但如此，新概念本身只是在它们加以解决的那些问题的基础上，只是在把它们包括在内的那些定理的基础上才能发生、发展、精确化和概括化. 变量和函数的概念不是一下子就现成地从伽利略、笛卡儿、牛顿或任何人那里产生出来的. 它们在许多数学家那里萌芽（例如，在纳皮尔那里与对数联系着），然后在牛顿和莱布尼茨那里采取了或多或少清晰的，但还远不是最终的形式，以后

它们随着分析的发展而精确化和概括化. 它们的现代定义直到十九世纪才形成,但是这种定义也不是绝对严格和完全终结了的. 函数概念本身的发展直到现在还在继续着.

数学分析是在已形成着的力学的材料基础上, 在几何问题和从代数引出的方法和问题的基础上建立起来的.

在变量数学的建立中第一个决定性步骤出现在1637年笛卡儿的作品《几何学》中,在这本书里奠定了所谓解析几何的基础,下面讲一讲笛卡儿的基本思想.

假设我们有方程:

$$x^2+y^2=a^2. \tag{6}$$

在代数中,把 x 和 y 理解为未知数,但是因为所给的方程不能把这两个未知数确定,所以从代数观点看来对它没有什么大的兴趣. 而笛卡儿不把它们看作应该从方程解出的未知数, 却把它们看作变量; 这时方程本身就表示这两个变量之间的依赖关系. 这种方程,在把所有的项都移到左边以后,可以写成如下的一般形式:

$$F(x, y)=0.$$

其次, 笛卡儿在平面上引进了现在称之为笛卡儿坐标 (图 6) 的 x, y, 于是每一对值 x, y 都对应于一个点, 反过来, 每一个点对应于它的坐标 x, y. 因此,方程 $F(x, y)=0$ 就决定了平面上的一些点的轨迹,而这些点的坐标满足这个方程. 一般说来,这是

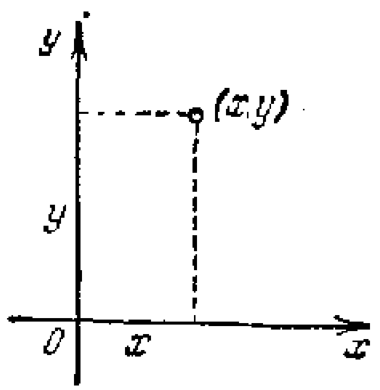


图 6

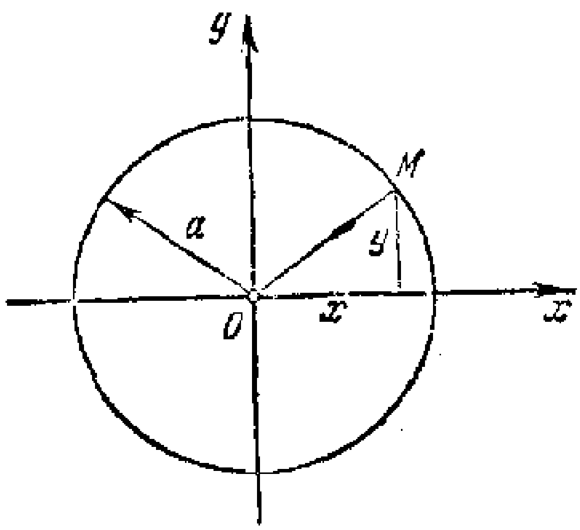


图 7

某一条曲线．例如，方程(6)决定一个以坐标原点为中心以 a 为半径的圆．事实上从图 7 可以看出：按照毕达哥拉斯定理，x^2+y^2 等于从原点到坐标为 x, y 的 M 点的距离的平方．所以方程 (6) 确定了与原点的距离是 a 的那些点的轨迹，也就是圆．

反之，由这样或那样几何条件所决定的点的轨迹也可以给出借助于坐标用代数语言表达同样条件的方程．例如，决定圆的几何条件——圆就是与给定点等距离的那些点的几何位置——这个条件若用代数语言即以方程(6)表示．

这样，解析几何的一般课题和方法就是：以带有两个变量的这个或那个方程来表示平面曲线并且根据方程的代数性质来研究相应曲线的几何性质，以及反过来：根据给出曲线的几何条件，找出它的方程，然后再根据方程的代数性质来研究这条曲线的几何性质．几何问题可以用这样的方法归结为代数问题，以及归根到底归结为数量的关系和计算．

关于解析几何方法的内容将在第三章中详细说明．现在我们只注意到这样一点，即从我们以上简短的叙述中就可以看出：几何、代数和一般变量概念的结合是这个方法的起源．解析几何基础中的主要几何内容是圆锥曲线(椭圆、双曲线、抛物线)的理论．这个理论，我们已经指出过，早在古代就得到了发展；阿波洛尼的结论已在几何形式上包括了圆锥曲线的方程．这种几何内容同希腊时期以后由数学发展已准备好的代数形式结合，又同由研究运动而产生的一般变量概念结合就构成了解析几何．

如果对希腊人来说，圆锥曲线只是具有纯粹数学兴趣的对象，那末到了笛卡儿时代，研究它们就对天文学、力学和技术都有现实的意义．刻普勒 (1571—1630) 发现行星是沿椭圆形轨道围绕太阳运动，伽利略则确定：投掷的物体，比如石头或炮弹，是沿抛物线飞出(如果可以忽略空气阻力，则是第一级的近似)．因而，有关圆锥曲线各种数据的计算成为迫切必要的了．笛卡儿的方法正是解决了这个迫切的问题．总之，这个方法是由以前数学的发展所准备好的，是由科学和技术的迫切需要而引起的．

3. 牛顿和莱布尼茨在十七世纪后半叶建立了微积分，这是变量数学发展的第二个决定性步骤. 这也正是分析的实际产生，因为微积分和解析几何不同，它的对象是函数本身的性质，而解析几何的对象毕竟还是几何图形. 事实上牛顿和莱布尼茨只是把许多数学家都曾参加过的巨大准备工作完成了，它的原理却要溯源于古代希腊人所创制的确定面积和体积的方法.

我们不在这里解释微积分基本概念的内容，以及继它们之后的分析理论——这些理论将在专门一章中说明. 我们现在只注意微积分的起源，主要是力学的一些新问题和几何的一些相当老的问题：作曲线的切线和确定面积和体积等问题成了微积分的起源. 还在古代就有人研究过这些问题（只要提到阿基米德就足够了）；在十七世纪初刻普勒、卡瓦列里和其他许多数学家也研究过这些问题. 但是这两类问题之间的显著关系的发现以及解决这些问题的一般方法的形成具有决定性的意义——这应归功于牛顿和莱布尼茨.

从坐标方法中引出表示一个量对另一个量的依赖关系的图示法（即函数的图示法）的可能性是发现力学问题和几何问题之间的显著关系的基础. 根据以曲线表示函数的图示法，容易阐明作为微积分起源的力学问题和几何问题的关系是什么，以及微积分的内容是什么.

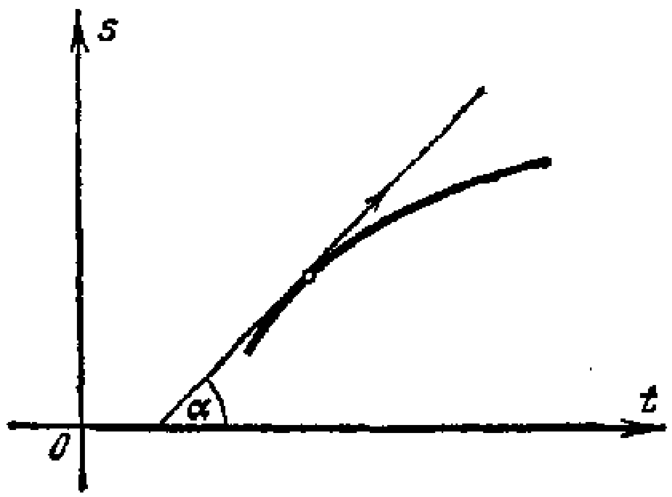


图 8

已知路程对时间的关系时，微分基本上就是寻求运动在任意给定时刻的速度的方法. 这个问题是用“微分法”解决的. 而这个问题和在给出路程与时间关系的图示的那条曲线上作切线的问题完全相当. 在 t 时刻的速度就是图上对应于 t 的那一点上的切线的倾角的正切

值(图 8).

已知速度对时间的关系时，积分基本上就是寻求所通过的路程的方法(或者一般地,寻求变量作用的总结果)．这个问题,显然是微分问题即寻求速度问题的反问题，它是用“积分法”解决的．而这个问题又和寻求在给出速度与时间关系的图示的那条曲线下的面积的问题完全相当．从 t_1 时刻到 t_2 时刻所通过的路程等于图上速度曲线下，对应于 t_1 和 t_2 的两条直线之间的面积值(图 9)．

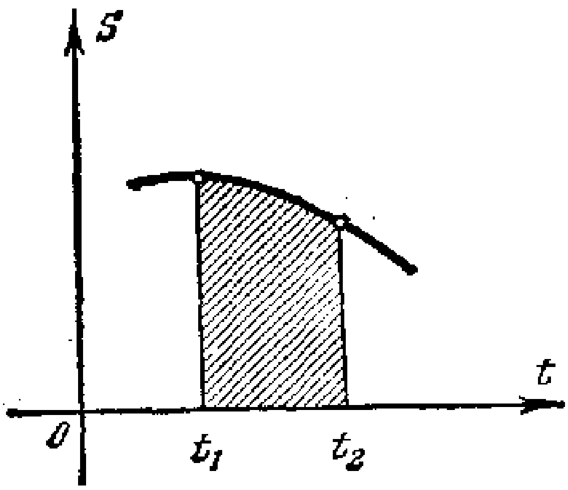


图 9

为了得出纯粹形式的微积分问题的一般概念，应该摆脱开对微积分问题的力学方式的提法,应该讨论一般函数,而不是讨论路程与时间或速度与时间的依赖关系.

除了变量和函数概念以外，以后形成的极根概念也是微积分以及进一步发展的整个分析的基础． 在分析形成的时期中，极限概念代替了当时运用的不大确切的无穷小量概念．按照路程变化规律,实际计算速度——“微分”和按照速度实际计算路程——“积分”的方法是以结合极根概念运用代数为基础的． 分析就是由于这些概念、方法与我们提到过的各种力学问题、几何问题以及某些其他问题(例如，求极大和极小的问题)相结合而产生的． 分析对于力学的发展已是必不可少的东西，在力学规律自身的形成过程中就已隐讳地出现了分析的概念．牛顿自己把他的第二定律表述成:“动量的变化和作用力成正比”．更确切些说:动量变化的速度和力成正比． 为了运用这条定律，就必须会确定某个量的变化速度，就是说会求微分． 如果我们用加速度与力成正比的说法来表述这条定律，那末问题仍然存在，因为加速度无非是速度变化的速度．同样地容易理解:为了确定由给定的变化着的力引起的，也即

一般常说依给定的变加速度进行的运动的规律，就必须会解决相反的问题——由量的变化速度找出这个量本身，就是说，必须求积分。可以说，牛顿为了有可能发展力学才被迫发明了微积分。

4．同微积分一道，还产生了分析的另外一些部分：级数理论(见第二章§14)，微分方程论(见第二卷第五章，第六章)，将分析应用于几何以后，就从几何中分离出一个特殊部门——称为微分几何，是关于曲线和曲面的一般理论(见第二卷第七章)．所有这些理论也都是因力学、物理学和技术问题的需要而产生并向前发展的．

微分方程论——分析的最重要的一支，是研究这样一种方程，方程中的未知项已经不是量，而是函数，即一个量对另一个量或另几个量的相依规律．容易理解，这样的问题从何而来．在力学中要求确定物体在给定条件下的运动规律，而不是速度或路程的某一个值．在流体力学中，要求按照流体总量找出速度的分布，就是找出速度对于空间三个坐标以及时间的依赖关系．同样，在电磁理论中要求找出整个空间中的场强，就是场强对于三个坐标的依赖关系．还有许多其他类似的问题．

这类问题常常是从包括流体动力学和弹性理论的力学、声学、电磁理论以及热学中产生的．总之，从分析本身产生之日起，它的发展就紧密地联系着力学和整个物理学的发展．分析的一些最伟大成就总是和解决上述科学部门所提出的问题有关．从牛顿起，最伟大的分析学家：伯努利(1700—1782)和欧拉(1707—1783)，拉格朗日(1736—1813)和庞加来(1854—1912)，奥斯特洛格拉特斯基(1801—1861)和李雅普诺夫(1857—1918)以及许多其他的人，他们为分析开辟新道路的工作通常都是从当时精确科学的迫切问题出发的．

新理论就是这样产生的：欧拉和拉格朗日和力学直接相联系建立了分析的新分支——所谓变分学(见第二卷第八章)，而在十九世纪末，庞加来和李雅普诺夫又是从力学问题出发，建立了所谓微分方程定性理论(见第二卷第五章§7)．

在十九世纪，新的重要分支——复变函数论使分析的内容更充实了(见第二卷第九章)．这个理论的萌芽曾出现在欧拉和某些其他数学家的作品中，但是直到十九世纪中叶，而且很大程度上是从法国数学家柯西(1789—1857)开始才形成严谨的理论．由于这个理论本身内容的丰富，更因为它能够深刻地渗透到分析的一系列定律中，并且它对于数学本身，物理学以及技术的许多重要问题的解决都有重要的应用，所以它很快地就得到极大发展，获得重大的意义．

分析蓬勃地发展着，它不仅成为数学的中心和主要部分，而且还渗入数学中较古老的范围，如代数、几何甚至数论．人们开始把代数理解为基本上是关于以多项式来表示的单变量或多变量函数的学说[1)]．解析几何和微分几何开始在几何中居于统治地位．最后，欧拉已经把分析方法引入数论，从而奠定了所谓解析数论的基础，它的发展联系着关于整数的科学的最深刻成就．

通过分析及其变量、函数和极限等概念，运动、变化等思想，使辩证法思想渗入了全部数学．同样地，基本上是通过分析，数学才经受了精确科学和技术的影响，数学才在自然科学和技术的发展中成为精确表述它们的规律和解决它们的问题的方法．正像在希腊人那里，数学基本上就是几何一样，可以说，在牛顿以后，数学基本上就是分析了．当然，分析不能包括数学全部；在几何、数论和代数中都保留着它们特有的问题和方法．比如，还在十七世纪，与解析几何同时，就产生了几何的另外一支——射影几何，纯粹几何的方法在射影几何中占统治地位．它以物体在平面上的映像(射影)问题作为自身的起源，因而特别应用于画法几何．

这时还产生了数学的重要新部门——概率论，它以在大量现象中发现的规律性作为自己的对象，例如许多次的射击和投掷钱

---

1) 这就是，例如 $y=a_0x^n+a_1x^{n-1}+\cdots+a_n$ 这样形式的函数．这个时期代数的基本问题(解方程 $a_0x^n+a_1x^{n-1}+\cdots+a_n=0$)无非就是要找出这样一些 x 的值，把这些值代入后函数 $y=a_0x^n+a_1x^{n-1}+\cdots+a_n$ 就等于零．解(方程的根)本身的存在，即代数的基本定理，是用分析来证明的(见第四章第三节)．

币. 概率论近来对物理学和技术具有特殊的意义；俄国的和苏联的数学家们的工作在促使概率论的兴盛方面起了很大作用，它的兴盛也决定于从自然科学和实践中提出的问题以及分析方法的运用. 这个理论的特点是：它研究"随机事件"的规律，给出了研究出现于偶然性中的必然性的数学方法. 概率论的基础将在第二卷第十一章中说明.

5. 尽管分析具有抽象的性质，却给予自然科学和技术以解决各式各样问题的有力方法. 我们已经提过其中的一些问题：已知某量本身和时间的关系时求这量的变化速度；确定曲线图形的面积和物体的体积；确定某个过程的总结果或变量的总作用. 又如，积分可以确定当压力按已知规律变化时气体膨胀所作的功；积分还可以从确立一个点电荷的场强的库伦定律出发，计算任何复杂的电荷系统的场强，等等.

此外，分析给出在这样或那样条件下求量的最大值和最小值的方法. 例如，借助于分析，容易确定当体积一定时具有最小表面积的圆柱形水槽的形状，因而耗费材料最少. 可以证明当水槽的高等于它的底的直径时就是所要求的形状（图 10）. 分析可以找出这样一条曲线的形状，物体沿着它从一个给定点滚到另一个给定点所需时间最短（这条线就是所谓的旋轮线，图 11）.

读者将在第二章和第二卷第八章中知道怎样来解决这些以及另外一些类似的问题.

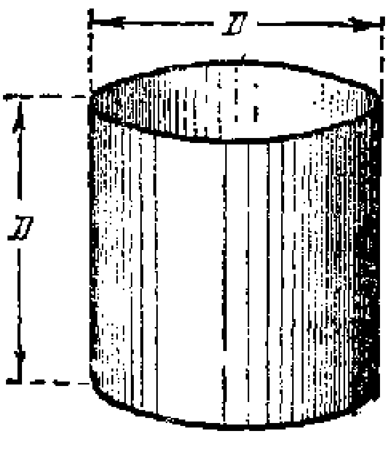


图 10

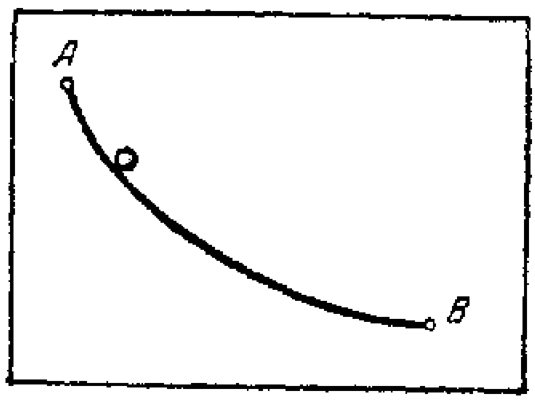


图 11

分析，更准确地说是微分方程论，不仅给出寻找变量的个别值的可能性，而且给出寻找未知函数也即一些量对另一些量的依赖规律的可能性．比如，我们能够根据电流的一般规律，计算把电压接到带有已知电阻、电容和自感的电路中时电流强度和时间的关系．我们能够确定液体流动的规律，以及在给定流动条件下，整个流体的速度分布规律．我们能够导出弦和膜的振动的一般规律，在各种介质中振动传播的规律：包括声波、电磁波、以及地震和爆炸发生时在地中传播的弹性波等；顺便提一下，这就对勘探有价值的矿藏和深入研究土壤提供了新的方法．读者将在第二卷第五章和第六章中碰到几个这类的问题．

最后，分析不仅给出这些或那些问题的解决方法，它还给出了精确科学的定量规律的数学表述的一般方法．正如我们以前已经讲过的，力学的一般规律如果不用分析概念就不能在数学上表述出来，而没有这样的表述，我们就没有可能解决力学问题．同样地，热传导，扩散，振动的传播，化学反应的进程的一般规律，电磁学的基本规律以及许多许多其他规律如果不借分析概念简直都不可能在数学上准确地表述出来．只是由于有了这样的表述，这些规律才能用于各种具体情况，才能在涉及热传导、振动、溶解、电磁场的各个问题中，在力学、天文学、物理学的众多的部门中，在化学、热工、力能学、机器制造、电工等等诸如此类的问题中得到准确的数学结论．

6．在希腊几何的历史上，欧几里得所作的严格和系统的叙述结束了以前发展的漫长道路．和这情况相似，随着分析的发展也必然要求对它作出论据，比它的实际方法的第一批创造者：牛顿、欧拉、拉格朗日和其他人所作出来的要更严格和系统化．他们所建立的分析随着自己的成长，首先导致越来越深刻和困难的问题，其次它的内容就要求它的基础具有很大的系统性和周密性．这样，理论的数量上的增长必然引起更好地论证理论，使理论系统化，批判地审查理论的基础等这样一些任务．理论的论证表现为理论的现有发展的总结，而不是出发点，因为离开理论，简直不知道应该

论证什么．恩格斯说："原则不是研究的出发点，而是它的終了的结果．"[1] 顺便说一下，现代某些形式主义者们常常忘记这一点，以为从公理出发来说明甚至发展理论是最合适的，而不需对这些公理所应当总结的现实内容作任何甄别．但是公理本身需要具有内容的论证；它们只不过总结了另外一些材料并对理论的逻辑结构提供基础[2]．

对于分析来说，批判，系统化和论证的必要时期是在上世纪中叶来到的．这项重要和困难的工作由于许多杰出学者的努力而胜利地完成了．特别是获得了实数、变量、函数、极根、连续性等基本概念的严格定义．

但是，正如我们已经指出过的，这些定义的任一个都不能认为是绝对严格和完全終结了的．这些概念正在继续发展．欧几里得和他以后两千年之内的所有数学家，毫无疑问，都认为欧几里得的《原本》一书几乎是逻辑严格性的标准．但是现在，从现代的观点看起来，欧几里得对几何的论证是十分表面的．这个历史的例子告诉我们，不应该迷惑于对现代数学的"绝对"和"彻底"的严格性的估计．在还没死去和变成木乃伊的科学中，没有也不可能有什么完全终止了的东西．但是我们可以确信地说：第一，现在已经确立的分析基础能够很好地适应现代科学任务，适应关于逻辑准确性的现代概念；第二，这些概念的继续深化和围绕这些概念正在进行着的讨论没有引起也不会引起人们简单地抛弃这些基础：但是却将导向对这些概念的新的、更准确和更深刻的理解，至于其结果，现在也许还难于完全判断．

虽然理论原则的建立是其发展的总结，但不能成为它的终结，相反地，正是它的新的运动．分析的情形也是这样．由于它的基础的准确化产生了新的数学理论——在上世纪七十年代由德国数学家康托尔所建立的任何抽象对象的无穷集合的一般理论，比如

1) 恩格斯，反杜林论，43 页，人民出版社，1956 年版．

2) 公理的这种二重性的作用，有时甚至在带有方法论性质的文集中都遭到忽视，从而把不是它的固有的作为理论的绝对论证的意义硬加给公理结构．

数的集合，点的集合，函数的集合或另外一类“东西”的集合．在这些思想的基础上成长起分析的新部门——所谓实变函数理论——关于它的概念，以及关于分析的基础和集合论的概念都将在第三卷第十五章中说明．同时，集合论的一般思想渗入到数学的所有部门．但是这种“集合论观点”与数学发展的新阶段不可分割地联系着，我们现在就转到对这个新阶段的简要考察．

§7．现代数学

1．数学教育上的各个阶段很自然地与§5中讲过的数学发展的四个阶段相适应，在不同教育阶段上所得到的数学知识水平，足够准确地相当于每一发展阶段的基本内容．

在数学发展的第一个时期所获得的算术和几何的基本结果已为我们大家所熟知，并且成为初等教育的内容．例如，为了确定完成某件工作，如铺装地板所需用的材料的数量，我们已经运用了数学的这些最初结果．

第二个时期——初等数学时期的最重要成就已成为中等学校的教学内容．

第三个时期的基本结果(分析的基础，微分方程论，高等代数等)成为每个工程师的数学训练的主要内容；除了人文科学部门以外，每一个高等学校，每一个系都要这样或那样地学习这些内容．因此，这个时期的数学的基本思想和结论已广泛地为众人所知，几乎所有的工程师和自然科学工作者都或多或少地运用着这些结果．

相反，数学发展的最近阶段，现代阶段的思想和结果基本上只是在各专业的物理-数学系学习．除了数学方面的专家，还有力学、物理学以及一些新技术部门领域中的科学工作者运用这些思想和结果．当然，这完全不表示它们与应用脱节了，但是它们是科学发展中的最新结果，自然显得更复杂些．所以现在转到阐述数学发展最近阶段的一般特征时，不能指望我们简短地讲述的全部内容

都完全明白易懂．我们试图简略地仅仅给出数学的这些新分支的最一般的特征，它们的内容将在本书相应的各章中更详尽地加以解释．

如果这一节读起来过于困难，初读时可以略过．而在读过后面各专门章节以后，再回头来读这一节．

2. 数学发展的现代阶段的开端，以其所有基础部门：代数、几何、分析中的深刻变化为特征．以几何为例加以考察，这种变化就可以看得最为清楚．新的非欧几里得几何学在1826年由罗巴切夫斯基几乎也同时由匈牙利数学家雅诺什·波约所发展．罗巴切夫斯基的思想远不是一下子就为所有数学家所理解的，因为这些思想太大胆和出乎意外了．但是，正是从这时候起，开始了几何的原则上的新发展，改变了对于几何是什么的本来理解．几何的应用对象和范围很快地扩大了．1854年著名的德国数学家黎曼继罗巴切夫斯基之后在这个方向上完成了最重要的步骤．他明确地表述了几何所能研究的“空间”数目无限的一般思想，同时指出了这种“空间”的可能的现实意义．

在几何的新发展中以下列两个情况为特征．

第一，如果在以前，几何只是研究物质世界的空间形式和关系（并且仅仅在欧几里得几何范围所反映的程度上），那末在现在，现实界的只是同空间形式和关系相似的其他形式和关系也成了几何的对象，所以在研究它们的时候，可以利用几何的各种方法．因此，“空间”这个术语在数学中获得了新的更广泛的，也是更专门的意义，同时几何方法本身也大大地丰富和多样化了（反过来，这些方法提供了认识我们周围物理空间的更完善的工具，几何的最初形式就是从这个物理空间抽象出来的）．

第二，甚至在欧几里得几何中也发生了许多重要变动：其中研究了复杂得多的图形乃至任意点集的性质．出现了对所研究的图形性质本身的原则上的新观点．把性质划分为单独的各组．这些性质是脱离了其他性质来加以研究的，这种脱离，这种抽象已经在几何内部形成一些独特的几何分支，这些分支本质上都是独立的

"几何". 几何的发展在所有这些方向上继续着，各种新而又新的"空间"和它们的"几何"：罗巴切夫斯基空间，射影空间，各种不同维数的欧几里得空间和其他空间，例如，四维的黎曼空间、芬斯勒空间，以及各种拓扑空间等，都成为几何研究的对象. 这些理论既在把几何除外的数学本身中具有重要应用，在物理学和力学中也有重要的应用，并且特别值得注意的是它们在作为空间、时间和引力的现代物理理论的相对论中的应用. 从以上所讲的可以看出，我们谈到的是几何的质变.

现代几何的思想以及几何中所研究的各种空间的学说的某些要素将在第三卷第十七章和第十八章中加以叙述.

8. 代数也发生了质的变化. 在上一世纪的前半叶产生了一些新理论，这些新理论引起了代数的变化，引起了代数的对象和应用范围的扩展.

正如我们在 § 5 中所讲过的，代数在它最初的基础上是关于对数字的算术运算的学说，这种算术运算是脱离了给定的具体数字在一般形态上形式地加以考察的. 这种抽象表现为下列事实，即代数中，凡量都以字母表示，按照一定的形式法则对这些字母进行计算.

现代代数在保持这种基础的同时，又把它大大地推广了. 现在在代数中还考察比数具有更普遍得多的性质的"量"，并且研究对这些量的运算，这些运算在某种程度上按其形式的性质说来是与加、减、乘、除等普通算术运算类似的. 矢量是最简单的例子，我们知道，矢量按照平行四边形法则相加. 在现代代数中进行的推广达到这样的程度，以致"量"这个术语本身也常常失去意义，而一般地是讨论"对象"了，对于这种"对象"可以进行同普通代数运算相似的运算. 例如，两个相继进行的运动显然相当于某一个总的运动，一个公式的两种代数变换可以相当于一个总的变换等等. 与此相应，就可以谈到运动或变换所特有的一类"加法". 现代代数在一般抽象形式上研究所有这样一些和另外一些类似的运算.

向这个方向发展的新代数理论，是从上一世纪前半叶的许多

数学家的研究中形成的，其中尤以法国数学家伽罗华（1811--1832）著称．现代代数的概念、方法和结果在分析、几何、物理以及结晶学等等中都有重大的应用．特别是在§3末提到过的费得洛夫所发展的晶体对称学说就是依靠几何与新代数理论之一——所谓群论的结合．

我们看到，问题涉及的是代数的对象的根本的、质的推广，以及对于代数是什么的理解本身的变化(现代代数的思想和它的某些理论的要素将在第三卷第二十章和第十六章中叙述)．

4．具有其抽象性的分析也发生了深刻变化．首先，正象在上一节中所讲过的，它的基础得到了精确化，特别是得到了它的基本概念：函数、极限、积分，最后是变量概念本身的精确和普遍的定义(还给出实数的严格定义)．分析基础精确化的原则是由捷克数学家波尔察诺（1781—1848），法国数学家柯西（1789—1857）以及许多其他数学家建立的．这种精确化与代数和几何的新发展属于同一时期；它在相当程度上是在上一世纪八十年代由德国数学家魏尔斯特拉斯、戴德金和康托尔完成的．其次，正如§6末所讲的，在数学新思想的发展中起过重大作用的无穷集合理论奠定了基础．

与集合论有关的变量和函数概念的精确化为分析的进一步发展打下基础．转向了对更一般的函数的研究；在相应的方向上，分析工具——微分和积分得到了推广．比如，在我们这一世纪的前夕产生了已在§6中提过的叫作实变函数理论的分析新部门，这个理论的发展首先应归功于法国数学家波雷尔、勒贝格等人，然后应特别归功于鲁金（1883—1950）及其学派．分析的所有新部门整个地称之为现代分析，以便同以前的、称为经典的分析相区别．

在分析中还产生了其他的新理论．例如分离出一个函数逼近理论的特殊部门，这是研究一般函数以各种“简单”函数，首先是以多项式即形如

$$a_0x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n$$

的函数来最好地近似表示的问题．

函数逼近理论仅仅在为函数的实际计算，为较简单的函数近似地代替复杂函数提供一般根据方面，就有着极重要的意义．这个理论的萌芽还是在分析产生的同时就发生了．伟大俄国数学家车比雪夫（1821—1894）给它指出了新方向．这个方向以后在所谓函数结构理论中得到发展，主要是由于苏联数学家的工作，特别是伯恩斯坦（1880 年生）在这方面作出了最重要的结果．关于函数逼近问题将在第二卷第十二章讲得．

以前已经讲过复变函数理论的发展．我们应该还记得由庞加来（1854—1912）和李雅普诺夫（1857—1898）的工作奠下了基础的微分方程定性理论，在第二卷第五章中将给出它的概念以及积分方程的理论等等．这些理论对于力学、物理学和技术都有很大的实际意义．比如，微分方程定性理论能解决关于运动、机械工作、电振动系统等等的稳定性问题．在最一般的意义上说，过程的稳定性就是：当起始条件或其运行条件有微小变化时，在全部时间过程中整个秩序的变化也小得微不足道．这类问题的技术价值是无须加以说明的．

5． 在分析和数理物理发展的基础上同几何和代数的新思想相结合，产生了新的宽广领域——在现代数学中起着特殊重要作用的所谓泛函分析．许多学者参与了它的创建工作；我们要提到，例如，近代伟大德国数学家希尔伯特（1862—1943），匈牙利数学家李斯（1880—1956），和波兰数学家巴拿赫（1892—1945）．年青的苏联学者获得了这个领域内与数理物理联系着的许多重要结果．将用第三卷第十九章单独一章来讲泛函分析．

这个数学新部门的本质可以简述于下．如果在经典分析中的量即“数”是变的，那么在泛函分析中就把函数本身看作是变的．某一给定函数的性质在这里不能独自地确定，而是在这个函数对另外一些函数的关系上确定的．所以考察的已经不是一些单个的函数，而是所有以这种或那种共同性质为特征的函数的集合，例如，所有的连续函数．函数的这种集合结合为所谓的“函数空间”．这和那些情形相对应，例如，我们可以考察平面上所有曲线的集合或

一定力学系统的所有可能运动的集合，在单个曲线或运动同其他曲线或运动的关系上来确定单个曲线或运动的性质.

从对单个函数的研究或探讨过渡到对变函数的考察类似于从未知数 x, y 过渡到变数 x, y, 这就是说类似于上节所讲过的笛卡儿的思想. 由于这种思想，笛卡儿提出了代数与几何的著名的结合——曲线方程，这种结合成为分析产生的决定性因素之一. 和这种情况相似，现在变函数的概念与现代代数和现代几何的思想相结合产生了新的泛函分析. 正如分析为当时建立的力学的发展所必需一样，泛函分析给出了解决数理物理问题的许多新方法，并成为新的原子力学、量子力学的数学工具. 历史在一定的程度上重复着，但是是按照新的式样，在更高阶段上重复. 正象我们所讲的，泛函分析把分析、现代代数和几何的基本思想和方法结合起来，反过来也对它们的发展显示出影响. 从经典分析中引出的许多问题现在正是常常以泛函分析为工具而获得了新的一般的解答. 现代数学的最一般和抽象的思想都集中在这里，如同集中在一个焦点上一样，并给出许多实际的效果.

从这个简短的叙述中，从这个分析的新方向（实变函数理论，函数逼近理论，微分方程定性理论，积分方程理论，泛函分析）的简单列举中就可以了解，我们讲到的，的确是分析发展的重要新阶段.

6. 在任何时候，计算工具的技术水平对数学方法本身都显示出重大的影响. 但是我们所掌握的实行计算的工具直到最近以前都是有很大局限性的. 最简单的设备（算盘式的），各种对数表和对数尺，算术计算机，最后，分析计算机和自动算术计算机——这些就是二十世纪四十年代以前所有的基本计算工具. 这些工具多少保证较快地实行一些个别的运算（加、乘等等）. 但是，为了得出实际问题的数字结果，时常要求按照复杂的程序进行极大量的运算，而且运算的程序常常取决于计算过程中得出的结果，这种问题的解答看来实际上是得不到的或者由于费时过长而完全失去价值.

在最近十年中我们眼看到整个计算技术水平发生了根本变化．根据新的原则建立起来的现代计算机可以使计算具有非常快的速度，并且可以按照预先指定的十分灵活的程序自动地进行复杂的计算序列．与现代计算机的结构和意义有关的某些问题将在第十四章中讲述．

新的技术不仅使以前难以实现的研究工作成为可能，而且改变了对许多已知的数学结论的估价．它特别促进了近似方法的发展，这种方法能够运用基本运算的序列达到具有足够大的准确度的必需的数字结果．并且对于这些方法是从在相应的机器上加以实现是否便利的观点上来估价的．

数理逻辑和新的计算技术的发展紧密相关．它首先是从数学内部由于本身发生困难的需要中发展起来的，并以数学证明的分析作为自己的对象．作为数学的一部分，数理逻辑把一般逻辑中那些客观上可以形式化并且可以用数学方法来发展的部分包括到自身中来．

一方面，数理逻辑溯源于数学的起源和基础，另一方面它又和计算技术的最新课题紧密相连，自然，那种导致建立一个特定过程（这个过程可以得到任意高的准确度的近似结果）的证明同关于这个或那个结果存在与否的更为抽象的证明是本质上不相同的．

由于研究这样一类问题（这类问题一般显然可以用导致唯一结果的，完全确定的方法来掌握）的普遍性的限度，产生了一类特殊范围的问题．在这方面数理逻辑得到了许多深刻的结果，这些结果从一般认识论观点看来也是十分重要的．

可以毫不夸张地说，在这个与新计算技术的发展和数理逻辑的成就紧密相关的新时期，现代数学的特点就是：研究对象不仅仅是这样或那样的客体，而且还有用来研究这个客体的那些方法和形式；不仅仅是这些或那些课题，而且还有解决这些课题的可能的方法．

对以上所说的还要补充一点：在现代数学的整个时期中，一些较老的数学部门——数论、欧几里得几何、经典代数和分析、概率论

等都继续急骤地发展着，以新的原则、思想和结果丰富着自己. 例如，俄国和苏联的数学家车比雪夫、费得洛夫、维诺格拉朵夫以及其他人在数论和直观几何中提出了许多思想和结果. 概率论的广泛发展与统计物理的一些最重要规律和许多现代技术问题联系着.

7. 从刚才已考察的几何、代数和分析的发展中所表现出来的整个现代数学的最一般的特征是怎样的呢？

首先是数学对象的大大扩展，它的应用范围的大大扩展. 数学对象和应用范围的这种扩展同时也表示数学在质上和量上的成长，新的理论与强有力的数学方法出现了，它们能够解决那些以前完全不可能解决的问题. 并且数学对象的扩展的特点首先就是现代数学有意识地给自己提出了研究量的关系和空间形式的可能类型的任务.

现代数学的另一个特点——这就是新的概括性概念的建立，新的更高的抽象程度. 正是这个特点保证了数学的统一，尽管数学的分枝不断成长并且多种多样. 在那些互相离得很远的领域中，概括性概念和理论揭示出统一的和一般的东西. 这些概念也保证了数学的基本部门——几何、代数、分析的方法的有相当的一般性，广泛的应用以及深刻的相互渗透.

集合论观点的统治地位也是现代数学特点. 当然，这个观点本身正是因为总结了由以前的数学发展所积累起来的内容丰富的材料才获得意义的.

最后，对数学基础的更为深刻的分析，对数学概念的相互关系，单个理论的结构的分析，对数学证明和结论的方法本身的分析也是现代数学的特点之一. 没有这种对基础的分析，概括性的原则和理论本身就不能进一步完善和发展.

现代数学的决定性特点可以这样来看，即它的对象不仅是已给出的，而且也是可能的量的关系和形式. 在几何中研究的不仅是空间的关系和形式而且还研究和空间的关系和形式相似的、可能的关系和形式. 在代数中研究的是各种具有可能的运算规律的抽象对象系统. 在分析中，不仅量是变的，而且函数本身也成为变

的．所有这种或那种类型的函数，也就是变量之间的可能的相互关系构成函数空间．因此，简要地说，如果初等数学是常量数学，继其后的一个时期的数学是变量数学，那末，现代数学就是各种量之间的可能的，一般说是各种变化着的量的关系和相互联系的数学．当然，这个定义是不完全的，不过它能一般正确地标志出现代数学的特点，这个特点决定了现代数学与过去时代的数学的质的区别。

§8. 数学的本质

1．现在我们用所有已经论述的作基础转到关于数学的本质的一般结论．

恩格斯在《反杜林论》的一节中说明了数学的本质，我们在这里将要引用这个精辟的片断．

读者容易在恩格斯的说明里认出前面已经说过的，例如，关于算术和几何的结论；这是可以理解的，因为我们叙述了数学发生和发展的实际历史，并且是以辩证唯物主义为指导来理解这个历史的．辩证唯物主义所以能得出正确的结论，正是因为它不把任何东西强加于事实，而是把事实按它们本来的样子加以研究，也就是说，在它们的必然联系和发展中加以研究．

恩格斯对于数学本质的论述是从批判杜林的荒诞观点，特别是所谓数学研究的是同经验无关的“纯粹理性”的创造物这种错误见解开始的．恩格斯写道：

“可是如说在纯数学中理性所涉及的只是自身的创造和想象的产物，那是完全不对的．数和形的概念不是从任何地方得来，而仅仅是从现实世界中得来的．人们用十个指头算数目，就是说作第一次的算术运算，这十个指头可以是一切别的东西，但总不是理性的自由创造物．要作计算不但要有被计算的对象，而且还要具有这样的能力，使其在考察这些对象时，能够摆脱其他的特性而仅仅顾到数目．而这种能力则是长期的依据于经验之上的历史发展

的结果. 和数的概念一样, 形的概念也完全是从外部世界得来的, 而不是在头脑中从纯粹的思维中产生出来的. 要能达到形的概念, 先应当存在具有一定形状的物体, 而且应把这些形状拿来比较. 纯数学是以现实世界的空间的形式和数量的关系——这是非常现实的材料——为对象的. 这些材料表现于非常抽象的形式之中, 这一事实只能表面地掩盖它的来自现实世界的根源. 可是为要能够在其纯粹状态中去研究这些形式和关系, 那末就必须完全使它们脱离其内容, 把内容放置一边作为不相干的东西; 这样我们就得到没有面积的点, 没有厚度和宽度的线, 各种的 a 和 b, x 和 y, 常数和变数; 只有在结末我们才达到理性本身自由创造和想象的产物——虚数值. 甚至在数学上的量的相互演算似乎是先验的, 也并不证明它们的先验的来源, 而只是证明它们的合理的相互关系. 矩形以其一边为中心而旋转得到的圆柱形, 要得到这样的概念, 那末先就需要研究一定数量的现实的矩形和圆柱形, 虽然是形式极不完全的矩形和圆柱形. 和其他所有科学一样, 数学是从人们的实际需要上产生的: 是从丈量地段面积和衡量器物容积, 从计算时间, 从制造工作中产生的. 可是和所有其他的思维领域一样, 从现实中抽象出来的规律, 在一定的发展阶段上就和现实世界相脱离, 并且作为某种好似独立的东西, 好似从外面来的规律——世界应当与此规律相适合——而与之相对立. 在社会和国家如此, 在纯数学上也正是如此, 而不是别的样子, 它也是往后被应用于世界上来的, 虽然它是从这一世界得出来的, 并且仅仅反映世界联系形式的一部分——仅仅因为如此, 数学才能被一般地应用"[1].

2. 这样, 恩格斯强调指出, 数学是反映现实界的, 它产生于人们的实际需要, 它的初始概念和原理的建立是以经验为基础的长期历史发展的结果. 我们已经在算术和几何的例子中十分详尽地研究了这个事实.

我们相信, 数、量、几何图形等概念正是这样产生的, 它们反映现实界的量的关系和空间形式. 分析的基本概念也同样是反映现

1) 恩格斯, 反杜林论, 37 页, 人民出版社, 1956 年版.

实的量的关系，它们是在概括大量具体材料的基础上逐渐建立起来的；例如，函数概念是在概括的、抽象的形式中反映现实的量之间的不同依存关系的。

综合了所有这些事实，恩格斯就得出这样的基本结论：数学以确定的完全现实的材料作为自己的对象，不过它考察对象时完全舍弃其具体内容和质的特点。这就使得，正如我们看到的那样，数学不同于自然科学，而恩格斯是明确地把数学从自然科学中划分出来的[1]。

数学之所以能够这样抽象地考察其对象是在其对象本身中有着客观根据的。那些反映在数学中的不依赖于质的特点或具体内容的一般形式、关系、相互联系和规律是不依赖于我们的意识而客观存在的。只有作为物体集合的客观性质的数的存在，数之间的相互关系之不依赖于对象的质的特点，以及这些相互关系之十分丰富才使算术成为可能。因此，如果那里没有这种不管内容的一般形式和关系，那里就不可能有数学的考察。

8. 数学的上述基本特点规定了它的其他一些特征。在§2中我们特别以算术为例考察了其中的某些特征。这些就是：独特的“公式语言”，应用的广泛，数学结论的脱离实验的特征以及它们的逻辑的不能避免和令人信服。数学的这种思辨的特征是它的非常本质的特点，我们现在来更详细地考察这个特点。

如果我们将数的概念从其具体根据中抽象出来，并且离开对这种或那种物体集合的任何关系来一般地考察整数，那末显然我们不能对这样抽象的数进行实验。始终保持这种抽象程度而不回到具体对象上去，那就只能通过从数的概念本身出发的推理得到关于数的新结论。当然，对于所有其他的数学结论也是这样。如果保持在纯几何学的范围内，就是说如果完全脱离任何质的、具体的内容来考察几何图形，那末除非用从这个或那个图形概念本身，从最基本的几何概念和公理本身出发的推理，我们就不能得到新的结论。例如，圆的性质是从圆之作为与一定点等远的诸点的轨

1) 恩格斯，反杜林论，10—11页，人民出版社，1956年版。

迹的概念中引伸出来的，完全不必考虑在实验中检验每个定论的问题。

这就是说，数学的抽象性质预先规定了这个事实，就是数学定理仅仅用从概念本身出发的推理来证明。

可以说，在数学中研究量的关系时，注意到的仅仅是它们的定义本身中所包含的东西．相应地，数学结论是用从定义出发的推理得到的．当然，仅从字面上来了解这些话并且认为数学概念的十分严格的定义的形成真正先于相应的数学理论的建立，那是不正确的；事实上，概念本身随着理论的发展，由于理论发展的结果而更加精确化，对整数概念的深刻分析，正如几何公理的精确公式化一样，不是在古代，而是直到十九世纪末才作出来．设想似乎有一种绝对精确地定义了的数学概念，那是更加错误的．任一概念，不管它是怎样被精确地定义了，也还是要变动的，它随着科学的发展而发展和精确化．这已经由全部数学概念的发展所证明，这只是再一次证实了辩证法的这样一条基本定理：世界上没有任何东西是完全不变和无论如何也不发展的．所以对于数学概念，第一，只能说到它们的充分的确定性，无论如何也不能说到它们的完全的确定性．第二，应该注意到它们的定义的精确性和明显性以及对它们分析的深度都是随着数学的发展而不断发展着．在下面一节中我们还有机会讲到数学概念的这种变动性，在这里我们只是记住这一点，而注意的却是数学概念的充分确定性．

正是数学概念的这种确定性以及逻辑本身的普遍意义成了数学结论具有内部确凿性和逻辑必然性这一特点的原因．思辨的数学结论的不可避免性给这样一种错误概念以借口，这种概念认为，似乎数学的基础是纯粹的思维，似乎数学是先验的，并非起源于经验，似乎它不反映现实界．例如，著名的德国哲学家康德就得出过这类的观点．这种极错误的唯心主义概念之所以产生，就特别是因为不在数学的现实的发生和发展中考察数学，而是在既成的形式上考察它．但是，这种方法是完全不行的，简单的理由就是：不符合事情的实际状况．数学不是先验的，而是从经验中产生的——

这是确凿不移的事实，顺便提一下，罗德的欧第姆早就写出了几何学的实际产生情况，我们在§3中已经引用过它。

不但数学概念本身，而且它的结论，它的方法都是反映现实界的。恩格斯正是揭示了这个重要情况，当他写道："数学的量似乎先验地一个自一个推导出来，这并没有证明它们的产生是先验的，而只是证明它们有合理的相互联系"。数学的结论和证明是作为人们在经验中所研究的各种现实联系的反映而产生的。数的加法反映了把几个物体集合现实地结合成为一个。关于两个三角形全等的定理的著名证明中讲到三角形的迭合，无疑地，这个证明是从把两个对象互相贴附在一起的实际操作中产生的，这种操作在比较对象的大小时经常采用。用积分计算体积是在抽象的形式中反映了用许多薄层迭加成一个物体或把一个物体切割成许多薄层的现实可能性。更复杂的数学证明是在这样的物质基础上进一步发展的结果。

4. 数学对象之完全舍弃掉任何具体性和以这为基础的数学结论的思辨特征，引起了数学的另一个重要特点：在数学中研究的不仅是直接从现实界抽象出来的量的关系和空间形式，而且还研究那些在数学内部以已经形成的数学概念和理论为基础定义出来的关系和形式。恩格斯正是注意到了数学的这个特点，当他说明点、线、常量和变量等概念的产生时说："……只有在结末我们才达到理性本身自由创造和想象的产物，这就是——达到虚数。"

虚数不是象我们说过的整数那样从现实界中提取出来的，这是一个历史事实。它们最初出现于数学内部，作为方程 $x^2=-a$ $(a>0)$ 的根，从代数的必然发展中出现的。后来虽然逐渐开始很随意地运用它们了，但是它们的现实意义仍然长久没弄清楚，这就是为什么要称它们为"虚数"的道理。以后，发现了它们的几何解释，它们得到许多重要的应用。同样地，罗巴切夫斯基的几何学也是作为这个伟大学者的创造物产生的；他还没有看到它的现实意义，所以称它为"想象的几何学"。但是它不是智慧的自由游戏，而是根据几何的基本概念作出来的必然结论，并且罗巴切夫斯基是

把它当作空间形式和关系的可能成立的理论来加以考察的．因此对于恩格斯所说的"自由创造和想象"，不能理解为简单的思想任性．科学中的自由创造——这是为来自经验的初始概念和原理所决定的有意识的逻辑必然性．

在由罗巴切夫斯基几何和精确虚数理论奠定基础的数学发展的新阶段上，产生了和不断产生着许多新的概念和理论，这些概念和理论是在已经形成的概念和理论的基础上建立的，而不是从现实界直接提取出来的．数学规定和研究现实界的各种可能形式，这正是数学发展的最近阶段的决定性特点之一．

辩证唯物主义的认识论对这个特点作了正确的理解．列宁说："认识是人对自然界的反映．但是，这并不是简单的、直接的、完全的反映，而是一系列的抽象过程，即概念、规律等等的构成，形成过程．"[1] 形而上学的唯物主义也承认：认识(其中有数学)是自然界的反映．但是，正如列宁所指出的，形而上学的唯物主义的根本缺陷就是不能把辩证法应用于反映论[2]．形而上学的唯物主义不了解这种反映的复杂性，不了解反映是通过一系列抽象过程进行的，是用在已经形成的概念和理论的基础上提出新概念，建立新理论的方法来进行的，是用不仅对经验中给出的东西加以考察而且也对可能的东西加以考察的方法来进行的．其实从已给出的东西到可能的东西这种过渡已经表现在象任何整数或者无穷长直线这样一些概念的形成中，因为在经验中既没有给出任意大的数，也没有给出无穷长的直线．但是当数的概念结晶出来以后，那末就从这概念本身中，就从用连续加一的方法形成顺序数的规律本身中，显示出数列无限延续的可能性．完全同样地，从沿直线行进中也显示出直线无限延续的可能性．这种可能性表现于欧几里得的第二公设中："可以无限地延长任一直线"．进一步抽象得到了关于整个自然数数列的概念和关于整个无穷直线的概念．在数学发展的最近阶段，经过一系列抽象过程和概念形成过程来建立理论，成为

1) 列宁，哲学笔记，167 页，人民出版社，1957 年版。

2) 列宁，哲学笔记，364 页，人民出版社，1957 年版。

具有新的性质的事情. 但是, 尽管抽象到这种程度, 数学仍然完全不能同现实界脱离. 新的数学理论在反映现实的基础上作为数学对象本身的逻辑结果而生长出来, 正是因为如此, 这些新的理论在运用于物理学和技术问题时又回到了现实界. 虚数就是这样的对于其他数学理论也是这样的, 不管这些理论怎样抽象.

各种多维空间的理论是一个具有特征性的例子. 它们在力学和物理学的影响下, 结合代数与分析的发展, 作为欧几里得几何学的推广而建立起来. 这些思想的结合引导黎曼去建立一般理论, 这个理论被其他数学家进一步发展, 得到了一系列重要的应用, 最后成为爱因斯坦创立广义相对论、更准确地说是引力理论的现成数学工具. 抽象的几何理论找到了这样光辉的应用不是偶然的, 不是由于"自然界与理性的前定的和谐", 而是由于它们本身是在直接从经验中产生的几何学的基础上成长起来的, 并且这些理论的建立者从这些理论产生的时候起就把它们同研究现实空间的任务联系了起来. 黎曼特别直接预见到自己的理论与引力理论的关系.

于是, 在数学的发展中体现了由列宁所表述的认识的运动规律: "当思维从具体的东西上升到抽象的东西时, 它不是离开——如果它是正确的……——真理, 而是接近真理. 物质的抽象, 自然规律的抽象, 价值的抽象及其他等等, 一句话, 那一切科学的(正确的、郑重的、不是荒唐的)抽象, 都更深刻、更正确、更完全地反映着自然. 从生动的直观到抽象的思维, 并从抽象的思维到实践, 这就是认识真理, 认识客观实在的辩证的途径."[1]

从以上所讲的可以明显地看出, 那种认为数学理论只是一套约定的公式, 用来在"思维经济原理"的基础上描述经验材料或者"排列感觉顺序"的唯心主义观点是完全错误的.

恩格斯指出 (见本节开始的引文), 从现实世界抽象出来的数学原理, 好象是作为某些现成的公式, 与现实世界相对立并被用来研究现实世界. 例如, 我们实际上经常把数字表为现成的形式加

1) 列宁, 哲学笔记, 155 页, 人民出版社, 1957 年版.

以利用. 对于在更高的抽象程度上产生的理论尤其是这样. 黎曼几何学对于引力理论是现成的数学公式, 就是已经提到过的一个例子. 但是恩格斯解释说: 应用数学来研究现实世界的这种可能性的根据在于: 数学从这个世界本身提取出来, 并且仅仅表现这个世界所固有的关系的形式部分, 因之才能够一般地加以应用. 许多理论是在数学本身内部建立的这个事实也丝毫不能改变这种解释. 这些理论作为现实的可能形式的理论而产生, 它们完全不是约定的, 因为它们是由于对象的逻辑发展本身的结果而必然产生的, 也正是因为如此它们才有许多现实的应用. 无论如何, 数学理论以这种或那种方式反映现实界, 所不同的只是在一些情形中反映比较直接, 而在另一些情形中反映要经过一系列抽象过程, 概念形成过程等等.

5. 数学发展的最近阶段的特征不仅是更高的抽象程度, 它的特征还在于数学对象的重大推广已经越出了对量的关系和空间形式的最初理解的范围.

在多维或无限维空间中的图形——这当然不是普通意义下的空间形式, 即当我们大家注意的只是普通的现实空间而不是数学的抽象空间时所理解的那种空间形式. 这些空间有现实的意义, 在抽象的形式上反映了现实界的一定形式, 但是这些形式只是与空间形式相似; 因此对于普通的现实空间来说可以称它们为"类空间". 在讲到多维空间和其中的图形时, 我们对空间概念加上了新的内容, 所以, 必须清楚地区别数学中广义的抽象空间概念和作为物质存在的普遍形式的原有意义下的空间概念.

产生在上世纪末、现在已得到广泛发展的新学科——数理逻辑, 可以作为数学对象越出量的关系和空间形式这些字眼的初始意义的范围的另一个例子. 数理逻辑考察的对象是数学结论的结构, 换句话说, 它研究那些命题可以用给定的方法从给定的前提中推导出来. 正如数学所具有的特性一样, 数理逻辑研究自己的对象是完全舍弃内容的, 因此以公式代替了命题, 而以运用这些公式的规则代替了论断的规则. 前提和结论之间的关系, 公理和定理

之间的关系当然不能归结为空间形式或普通意义下的量的关系，例如，归结为概念外延的关系。

我们指出群论作为另一个例子，群论可以理解为关于对称的最一般形式的学说。但是我们说，当硫从斜方形晶体过渡到角柱形晶体时，晶体对称的改变是物质状态的根本的质的改变。所以，群论是关于这样一些量或关于对象的这样一些规定性的学说，这些量或规定性的变化同对象自身的根本变化相伴随。

因此，数学对象的推广引起了量的关系和空间形式概念本身的重大推广。在这种情况下，这种推广了的数学对象的一般特征是怎样的呢？

如果不用列举的办法，而用力求表明那些存在于数学对象的全部多样性中的一般的和具有特征性的东西的办法来回答这个问题，那末本质上我们可以从恩格斯的著作中找到答案。只要不仅仅注意到他对数学对象的说明，而且注意到他对考察这种对象的方法的说明：把形式和关系完全从内容中抽象出来，就足以得到答案。数学的这个抽象的特点同时也给出了它的对象的定义。

数学的对象是现实界的这样一些形式和关系，这些形式和关系客观地具有与内容无关的性质，无关到这样的程度以致能够把它们完全从内容中抽象出来，并且能够在一般的形态中定义出来，达到这样的明确性和精确性，保持这样丰富的联系，以致成为理论的逻辑发展的根据。如果在一般说法的意义下也称这样一些关系和形式为量的关系和形式的话，那末可以简单地说，数学以纯粹形态的量的关系和形式作为自己的对象。

抽象绝对不是数学所特有的。但是其他科学感兴趣的首先是自己的抽象公式同某个完全确定的现象领域的对应问题、研究已经形成的概念系统对给定现象领域的运用界限问题和所采用的抽象系统的相应更换问题，并把这些作为最重要的任务之一。相反地，数学完全舍弃了具体现象去研究一般性质，在抽象的共性中考察这些抽象系统本身，而不管它们对个别具体现象的应用界限。可以说，数学抽象的这样绝对化才是数学所特有的。

正是数学所研究的形式具有上述与内容的客观的无关性规定了数学的基本特点：它的思辨的特点，它的结论的逻辑必然性和看来不可变动的性质，以及新概念和新理论产生于数学内部的事实；数学应用的特点被这个与内容无关的性质所规定．当我们能够把实际问题翻译成数学的语言时，我们同时能够撇开问题的次要的具体特点，利用一般公式和结论，得出一定的结果．所以，数学的抽象性成为了它的力量，这种抽象性实际上是必须的．

6．现在回到恩格斯关于数学的论断，我们可以看到，在这个论断中包含了多么深刻和丰富的内容和怎样的发展的可能性．他自己并不是数学家，却对这门科学的基础作出了这样深刻的分析，这不仅因为他是伟大的思想家，而最主要地，是因为他掌握了辩证唯物主义并以它为指导来解释数学本质的问题．不难理解，为什么在恩格斯以前没有人能够这样深刻和正确地解决这个问题．最伟大的数学家们也没有能够在这样的程度上解决这个问题．

同样地，以后列宁对物理学的问题给出了极为深刻的分析，这个分析超出了这个领域中所完成的一切工作．

这个事实再一次证明了辩证唯物主义的意义和力量；这个事实证明，为了掌握一门科学，认识它的个别原理是不够的，甚至在这门科学上作一个创造性的工作者也还不够——为了掌握一门科学，还必须掌握正确的一般方法，掌握辩证唯物主义．离开辩证唯物主义，科学结论或者成为轮廓模糊的一堆，或者表现为歪曲的形式；代替了对科学的正确认识而得出关于科学的荒诞的，形而上学的，唯心主义的观念．例如，许多不掌握辩证唯物主义的数学家或者完全不能了解自己这门科学的一般问题，或者对它们的解释是完全错误的[1]．

在恩格斯写《反杜林论》的时候，即在1876—1877年，非欧几

1）例如，指出这样一件事情是很有趣的：两个有名的美国几何学家魏布伦和怀特海德在自己的著作《微分几何基础》中，企图接触到几何学是什么的定义问题，并得出结论说，定义是不可能给出的，除非是这样："几何学是专家们称之为几何学的东西"．

何学和多维空间几何学刚刚在数学家之间得到承认，群论刚刚形成，集合论刚刚产生，而数理逻辑仅仅萌芽．所以可以理解，数学发展的新阶段的特点不能由恩格斯详尽地描述出来；但虽然如此，我们在他的论断中也可以找到对于理解这些特点的指示．

§9．数学发展的规律性

最后，我们试图简要地叙述一下数学发展的一般规律性．

1．数学不是任何一个历史时代，任何一个民族的产物；它是好几个时代的产物，许多世纪的人们工作的产物．我们已经看到，数学的最初的概念和原理在远古时代就产生了，还在两千多年以前就形成了严谨的体系．不管数学经历了多少次改造，它的概念和结论都仍旧保持下来，从一个时代过渡到另一个时代，例如，算术规则或毕达哥拉斯定理就是这样．

新的理论把先前的成就包括到自身中来，把先前的成就加以精确化，补充和推广．

同时，如同从上述的数学史的简要提纲中显然看到的那样，数学的发展不仅不能归结为新定理的简单积累，而且还包括了重要的、质的变化．与此相应，数学的发展划分为好几个时期，各个时期间的过渡正是以这门科学的对象本身或结构的这些根本变化为标志的．

数学把现实界量的关系的一切新领域包括到自己的范围之内来．同时，简单的，最直接的字面意义上的空间形式和量的关系仍然是数学的最重要的对象，而对新的联系和关系的数学理解由于已经形成的量的和空间的科学概念系统，并以这个概念系统为基础而不可避免地产生出来．

最后，数学本身内部成果的积累，必然要引起向新的抽象阶段，向新的概括的概念上升，以及对于基础和原始概念分析的深化．

正如一棵槲树在健壮的生长中，用新的树层使老枝变粗，长出

新枝，枝叶往上长高，根又往下长深一样，数学在自己的发展过程中把新的材料添加到已经形成的领域之中，形成新的方向，升到新的抽象高度，并在基础方面更加深化．

2．数学以现实界的现实形式和关系作为自己的对象，但是，正如恩格斯说过的那样，为了在纯粹形态上来研究这些形式和关系，必须把它们同它们的内容完全割裂开来，把它们的内容抛在一边当作没有什么关系的东西．但是，离开内容的形式和关系是不存在的，数学的形式和关系不能绝对地同内容无关．因此，数学按其本质企图实现这种割裂，是企图实现一种不可能的事情．这就是在数学的本质中的根本矛盾．这种矛盾是认识的普遍矛盾在数学方面的特殊表现．思想反映现实界的任何现象，任何方面，任何因素，都把这些现象、方面、因素粗糙化了，简单化了，把它们从自然界的普遍联系中割开了．当人们研究空间的性质，确定空间具有欧几里得几何的性质时，人们完成了一个极为重要的认识步骤，但是，在这之中就包含了一个错误：空间的现实的性质被简单化、公式化，抽象得离开物质了．但是，没有这个步骤就根本不可能有几何学，正是在这种抽象的基础上（既从它内部的研究中，又从数学的结果同其他科学的新材料的对照中），产生了和巩固了新的几何理论．

在日益接近于现实的各个认识阶段上不断解决和恢复上述的矛盾，就是认识发展的本质．并且，认识的积极内容，认识中绝对真理的因素当然是决定性的．认识依上升的路线前进，而不是踏步不前，和错误简单地混合在一起．认识运动就是不断克服其不精确性和局限性．

上述的基本矛盾随着引起了另一些矛盾．我们从不连续与连续的对立的例子中看到了这种矛盾．（在自然界中，不连续与连续之间没有绝对的分割，在数学中把它们加以分割不可避免地引起了创立更新的概念的必要性，这些新概念要更深刻地反映现实界，同时克服现存数学理论的内在不完善性．）有限和无限、抽象和具体、形式和内容等等全都一样，它们都是数学中的根本矛盾的表

现。但是，数学中的根本矛盾的决定性的表现却是在于：数学舍弃了具体的东西，周旋在自己的抽象概念的圈子中，因而脱离了实验和实践，可是同时，数学只有以实践为基础，只有不是纯粹的而是应用的数学时，才能是科学（即具有认识的价值）。用几句黑格尔式的话来说，纯粹数学不断地否定自己之作为纯粹数学；没有这种否定，它就不能具有科学的意义，不能发展，不能克服内部不可避免地产生着的困难。

数学理论在其形式上作为对于具体结论的某种公式同现实内容相对立。数学是把自然科学的量的规律化为公式的方法，是研究自然科学理论的工具，是解决自然科学和技术问题的手段。现阶段的纯粹数学的意义首先就在于数学的方法。正如一切方法不能自身存在和发展，只能在其应用的基础上，在同应用到它的内容相联系之中存在和发展一样，数学也不能离开应用而存在和发展。这里又表现出对立的统一：一般方法作为解决具体任务的手段，同具体任务相对立，但是它自身又是产生于具体材料的概括之中，只能在解决具体任务之中存在、发展和得到检证。

3．社会实践在数学的发展中从三个方面起了决定性的作用。社会实践向数学提出新的问题，刺激数学向这个或那个方向发展，并且提供检证数学结论的真理性的标准。

从数学分析的产生这个例子上可以非常清楚地看到这一点。首先，正是力学和技术的发展提出了从一般的形态上研究变量间的依赖关系的问题。阿基米德十分接近了微分和积分的计算，但却停留在静力学问题的范围之内，只是到了研究运动的时候，产生了变量和函数概念，才促使数学分析的形成。牛顿如果没有发展相应的数学方法，就不能发展力学。

其次，正是社会生产的需要促使所有这些问题的提出和解决。无论在古代社会，或是中世纪社会，这些推动力都还不存在。最后，非常值得注意的是，数学分析在其产生的时候正是在应用中来证实自己的结论。正是因为如此，它才能离开它的基本概念（变量、函数、极限）的严格定义而发展，这些严格定义是以后才给出

的．分析的真理性是由它在力学、物理学和技术中的应用而确立的．

在数学发展的所有时期中情形都是如此．从十七世纪开始，理论物理学和新技术问题同力学一道对数学的发展发生了最直接的影响．连续介质力学，后来还有场论(热传导、电、磁、重力场)引向了偏微分方程理论的发展．上世纪末开始的分子理论的研究和整个统计物理学，是概率论，特别是随机过程理论发展的重要推动力．相对论在黎曼几何及其分析方法和推广的发展中起了决定性的作用．

现在，新的数学理论如泛函分析等的发展，是由量子力学和电动力学问题、计算技术问题、物理学和技术的统计问题等等所推动的．物理学和技术不仅向数学提出新的问题，促使数学去研究新的对象，而且也促进了那些现在为它们所需要而原先在很大程度上是从数学本身内部形成的数学部门的发展，黎曼几何的情形便是这样的．简言之，为了迅速发展一门科学，不仅需要这门科学能解决新的问题，而且解决这些问题的必要性还应该是社会发展的需要所规定的．在数学中最近产生了许多理论，但是只有那些在自然科学和技术中找到了应用，或者在具有这种应用的那些理论的重要概括中起了作用的理论，才得到了发展，才能巩固地列入科学之中．而另一些理论却停滞不前，例如，某些没有找到重要的应用的精炼化了的几何理论(非戴扎尔、非阿基米德几何)就是这样．

数学结论的真理性不是在一般的定义和公理中，也不是在证明的形式的严格性中，而是在现实的应用中，也就是说，归根到底是在实践中得到最后的证实．

整个地说，数学的发展必须首先理解为其对象的逻辑(反映在数学本身的内在逻辑中)、生产的影响以及同自然科学的联系三者的相互作用的结果．这种发展是以对立的斗争的复杂途径进行的，其中包括数学的基本内容和形式的重要变化．从内容上来说，数学的发展决定于其对象，但是它基本上和归根到底是由生产的需要所促进的．数学发展的基本规律性就是如此．

当然，我们不应当忘记，这里说的只是基本的规律性，数学同生产的联系，一般说来，是复杂的．从前面所说的可以显然看出，如果企图论证每一种数学理论的出现都是直接的“生产定货”，那是幼稚的．而且，数学同一切科学一样，具有相对的独立性和自身内在的逻辑，这种内在逻辑，我们强调指出过，反映了客观的逻辑，即其对象的规律性．

4．数学一向不仅经受社会生产的最重要的影响，而且经受全部社会条件的整体的最重要的影响．它在古希腊的最发达的时代的辉煌进步，代数在意大利文艺复兴时代的成就，分析在英国革命之后的时代的发展，数学在法国紧接着法国革命的时代的进展——所有这些令人信服地显示了数学的进步同社会的一般技术、文化和政治的进步的不可分割的联系．

从数学在俄国发展的例子上也可以鲜明地看到这一点．从罗巴切夫斯基、奥斯特洛格拉斯基、车比雪夫开始的独立的俄国数学学派的建立，是不能同俄国整个社会的进步分开的．罗巴切夫斯基的时代——这是普希金、格林卡的时代，十二月党人的时代，数学的繁荣是普遍高涨的因素之一．

伟大的十月社会主义革命以后的时期中社会发展对数学的影响是更加令人信服的，在这段时期中，具有基本意义的研究在许多方向上以惊人的速度一个跟着一个出现，这些方向包括：集合论、拓扑学、数论、概率论、微分方程理论、泛函分析、代数、几何．

最后，数学一向经受而且现在仍然经受着思想体系的显著影响；如同在一切科学中一样，数学家和哲学家们是在这样或那样的思想体系的领域中领会和理解数学的客观内容的．

简言之，科学的客观内容总是被置放于这些或那些思想形式中；这两个辩证的对立面——客观内容和思想形式——的统一与斗争，在数学中，如同在一切科学中一样，对于这门科学的发展起了远非微末的作用．

与科学的客观内容相适应的唯物主义同与这一内容相矛盾并对它作歪曲理解的唯心主义之间的斗争，贯串在数学的全部历史

中．这个斗争还在古希腊时代就明显地表现出来，毕达哥拉斯、苏格拉底和柏拉图的唯心主义反对法勒斯、德莫克利特和其他建立了希腊数学的哲学家们的唯物主义．随着奴隶占有制的发展，社会的上层分子就脱离了生产，认为参加生产是下层阶级的命运，这就产生了“纯粹”科学同实践的脱离．只有纯粹的理论几何学才得到了真正的哲学家的应有注意．值得注意的是，柏拉图认为当时出现的某些力学曲线甚至圆锥曲线的研究是不属于几何学的范围之内的，因为这些研究“并不引导我们同永恒的非肉体的理想相往来”，而是“庸俗的手工业工具的运用中所需要的”．

数学中唯物主义反对唯心主义的斗争的一个鲜明的例子，是罗巴切夫斯基的活动，他提出和坚持了对数学的唯物主义理解来反对康德主义的唯心主义观点．

对于俄国数学学派，一般地是以唯物主义传统为特征的．例如，车比雪夫明确地强调实践的决定性作用，而李雅普诺夫则用下述的著名语句来表明我国数学学派的风格：“详尽地研究那些从应用的观点看来特别重要，同时又提出了特别的理论困难，要求发明新的方法和上升到科学原则的问题，然后就概括已经得到的结论，用这种方式建立或多或少一般的理论．”概括和抽象并不是本身存在，而是同具体材料联系着；定理和理论并不是本身存在，而是同归根到底要导向实践的科学有着普遍联系——实际上这正是很重要的和很有希望的[1]．

像高斯和黎曼这样的大科学家的意图也是这样的．

但是，随着资本主义在欧洲的发展，反映着十六世纪到十九世纪正在上升的资产阶级的进步思想体系的唯物主义观点，开始被唯心主义观点所代替．例如，康托尔（1846—1918）建立了无穷集合理论，却公然援引神说了这样意思的话，似乎无穷集合在神的理智中有着绝对的存在．十九世纪末二十世纪初法国的大数学家庞

1）一般说来，对数学各部门相互间的必然联系，以及与自然科学和实践的必然联系的理解不但对数学本身的正确观点有十分重大的意义，而且对学者们选定研究的方向和对象也有十分重大的意义。

加来提出了唯心主义的"约定论"观点，根据这种观点，数学是为了描述经验的多样性的方便而采用的有条件的约定系统．例如，按照庞加来的意见，欧几里得几何公理只不过是有条件的约定，它们的价值决定于方便和简单，而不决定于同现实界的一致．因此，庞加来说，例如，在物理学中宁可抛弃光直线传播的定律，而不愿抛弃欧几里得几何学．这个观点为相对论的发展所驳倒，相对论不顾欧几里得几何的全部"简单"与"方便"，却同罗巴切夫斯基和黎曼的唯物主义思想完全一致，得出结论说，空间的现实的几何学并不是欧几里得几何学．

在集合论中发生的困难的基础上，又由于对数学的基本概念进行分析的必要性，在十九世纪初，数学中间出现了不同的倾向．对数学的内容的理解的一致消失了；不同的数学家开始不仅按不同的方式考察这门科学的一般基础，这在以前也是有过的，而且甚至按不同的方式来评价一些个别的具体结果和证明的意义和价值．某些结论对一部分人来说被认为是有意义和有内容的，另一部分人却认为是没有意义和价值的．产生了"逻辑主义"、"直觉主义"、"形式主义"等唯心主义潮流．

逻辑主义者断言，全部数学是从逻辑概念中导出的．直觉主义者认为数学的来源在于直觉，认为只有能直觉地感受的东西才有意义．因此，他们特别完全否定康托尔的无穷集合理论的意义．此外，直觉主义者甚至否认像任何 n 次的代数方程都有 n 个根的定理这类判断的简单意义．在他们看来，这个判断是空洞的，如果没有指出算出根的方法的话．这样，完全否认数学的客观意义，就使得直觉主义者污辱了相当大部分的数学成就，说它们都是"没有意义的"．他们中最极端的人甚至认为，有多少数学家，就有多少种数学．

我们这个世纪初期的大数学家希尔伯特以他自己特有的方式来挽救数学免于这种攻击．他的思想的实质就是把数学理论归结为根据约定的规则对符号作纯粹形式的运算．他以为采取了这种完全形式的态度，一切困难都会消失，因为数学的对象只是符号以

及与它们的意义毫不相关的运算规则. 这就是数学中形式主义的提出. 用直觉主义者布劳埃尔的话来说, 在形式主义者看来数学的真理在纸面上; 而在直觉主义者看来, 数学的真理在数学家的头脑之中.

不难看出, 他们两方面都是不正确的, 因为数学, 连同写在纸面上的和数学家所想的, 都是反映现实界的, 而数学的真理则在于它同客观现实相一致. 所以这些思潮都把数学同物质的现实割裂开来, 都是唯心主义的.

希尔伯特的思想由于它自身的发展而遭到失败. 奥地利数学家哥德尔证明, 甚至算术都不能像希尔伯特所指望的那样完全形式化, 哥德尔的结论鲜明地揭示了数学的内在辩证法, 数学的任何一个领域都不能为形式的计算所穷尽. 甚至自然数序列的简单的无限性都是有限的符号及作用于符号的规则系统所不能穷尽的. 这样, 就从数学上证明了恩格斯就一般形式而说的这样一段话:

"无限性是矛盾……这种矛盾的消灭, 即成为无限性的终结."[1] 希尔伯特指望把数学的无限性归入有限系统的范围之内就可以消灭一切矛盾和困难. 这是不可能的.

但是, 在资本主义条件下, 约定论、直觉主义、形式主义和其他类似的思潮不仅保存着, 而且由对数学的唯心主义观点的一些新变种所补充着. 在某些主观唯心主义的新变种中, 严重地利用了与数学基础的逻辑分析有关的理论. 主观唯心主义现在之利用数学, 特别是数理逻辑, 并不下于利用物理学, 因此, 数学基础的理解问题具有特殊的尖锐性.

这样, 数学发展的困难在资本主义条件下产生了这门科学思想上的危机, 这个危机的基础与物理学的危机相似, 它的本质, 列宁在他的天才著作"唯物主义与经验批判主义"中已经说明了. 这个危机并不意味着数学在资本主义国家中完全停止发展, 许多站在明显的唯心主义立场上的学者, 在解决具体数学问题和发展新理论中, 作出了重要的, 有时是卓越的成就. 只要引证一下数理逻

1) 恩格斯, 反杜林论, 第 52 页, 人民出版社, 1956 年版.

辑的光辉研究就足够说明这一点了.

在资本主义国家中广泛流传的对数学的观点的根本毛病，就在于它的唯心主义和形而上学；在于使数学同现实界相脱离和忽视它的现实发展. 逻辑主义、直觉主义、形式主义和其他类似的思潮从数学中分离出某一个方面——同逻辑的联系，直觉的明显性，形式的严格性等等——没有根据地把它的意义加以夸大和绝对化，把它同实际割裂开来，深刻地分析了数学的某一个特征本身，而忘记了数学的整体. 正是由于这种片面性，这些思潮的任何一种，不管其个别结论如何精细和深刻，都不能达到对数学的正确理解. 同唯心主义和形而上学的各种潮流相对立，辩证唯物主义把数学以及整个科学看作它本来的样子，在其联系和发展的全部丰富性和复杂性中加以考察，正是因为辩证唯物主义企图理解科学同实际的联系的全部丰富性和复杂性，理解科学从简单的概括经验到高度抽象以及从抽象到实践的发展过程的全部复杂性，正因为辩证唯物主义经常引导自己对科学的看法去同科学的客观内容、同科学的新发现相适应，正因为如此，归根到底也只能是因为如此，辩证唯物主义就成了唯一真正科学的哲学，这种哲学导向对科学一般也包括对数学的正确理解.

文　　献

讨论数学的一般问题

A. H. 阔尔莫果洛夫，数学，苏联大百科全书第 26 卷，见高等教育出版社，1956 年版《数学·算术》一书.

A. Д. 亚历山大洛夫，几何学，苏联大百科全书第 10 卷，见《数学通报》1955 年第 4 期.

数学史的书籍

И. Г. Башмакова и А.П. Юшкевич, Происхождение систем счисления, Энцикл. элемент. математики, т. I, Гостехиздат, 1951.

Б. В. Гнеденко, Очерки по истории математики в России. Гостехиздат, 1946.

Г. Г. Цейтен, История математики в древности и в средние века. ГОНТИ, 1938.

Г. Г. Цейтен, История математики в XVI и XVII веках. ГОНТИ, 1938.
《三十年来的苏联数学 (1917—1947)》，阐明数学各部门在苏联的发展的论文选集。见切伯塔廖夫等著《代数学（三十年来的苏联数学）》一书，科学出版社，1954年版。

孙小礼 译

关肇直 校

第二章 数学分析

§1. 绪　　论

当中世纪晚期新的生产关系在欧洲出现，行将取封建制度而代之的资本主义萌芽的时候，伴随着而来的还有伟大的地理发现和探险．在1492年，根据大地如球的观念，哥伦布发现了新大陆．哥伦布的发现大大地扩展了当时所知道的世界范围并且引起了人类思想上的激变．在十五世纪末和十六世纪初，有伟大的艺术家与人文主义者里昂纳多·达·芬奇，拉斐尔，米开兰基罗在进行创作，于是革新了艺术．在1543年出版了使天文学完全改观的哥白尼著作《天体运行论》；在1609年出现了刻普勒的《新天文学》，其中载有关于行星绕日运动的第一与第二定律，而在1618年出现了他的《宇宙的和谐》一书，其中载有第三定律．伽利略从阿基米德著作的研究和大胆的实验出发，开始创立了对当时正在兴起的技术非常必需的新力学．在1609年，伽利略把他所自制的不大而且不完善的望远镜指向了夜间的天空．望远镜里的一瞥就已经足够粉碎亚里斯多德的理想天球和关于天体完美无缺的教条了．月亮的表面原来满布着山脉并且到处是环形山．金星显露出跟月亮一样的位相，而木星则有四个卫星环绕着，正好像太阳系的一个不大的直观模型．银河被分解成为个别的星体，因而第一次使人们理会到星体的惊人的遥远的距离．从来没有一件科学上的发现曾经对文化界起了这样的影响[1]．

远洋航行的发展和因此而引起的天文学研究，新技术的发展和与此有关的力学的发展都迫切需要寻求各种方法，来解决当时

1) 这一段是以瓦维洛夫院士的优美的"伽利略"一文(苏联大百科全书，第二版，第十卷，1952年；中译本：伽利略，人民出版社，1954年版)作为根据的．

所发生的许多新的数学问题．这些问题的新异之处主要在于必须对广义而言的运动从数量上来研究它的规律．

自然界是跟静止不动的状态格格不入的．像恩格斯所指出的那样，整个自然界，由其最小质点到最大物体，都是处在永恒的产生和消灭过程中，处在毫不间断的流动中，处在始终不停的运动和变化中．每一门自然科学所研究的，归根结底总是运动的某一个方面或另一个方面，运动的某一种形式或另一种形式．数学分析就是数学的这样一个分支，它对于各种变化过程、运动过程以及一个量与另一个量相倚而变的过程提供了进行数量上研究的方法．所以正在这样的时期——当技术与航海问题所促成的力学与天文学的发展已经使观察、测量和理论的假设充分地积累起来，紧接着便是引导科学对最简单形式的运动作数量研究的时候——有数学分析产生，并不是偶然的．

"无穷小分析"这一名称本身并不曾讲到研究的对象，却着重指出在数学的这一分支中所运用的方法．这是一种特殊的、数学上的无穷小法，在现代的形式中也就是极限的方法．我们立刻就要举出一些典型的例子来说明这种使用极限法的推论，而在后文的一节中还要把各项必要的概念一一明确化．

例 1 据伽利略用实验确定，真空中的自由落体在时间 t 内所经历的路程 s 由下列公式表出：

$$s=\frac{gt^2}{2} \tag{1}$$

（式中 g 是常量，等于 9.81 公尺/秒^2）[1]．问落体在其路程的每一点上速度有多快？

设落体在瞬时 t 经过某一点 A．让我们来看一看在随后一小段长达 Δt 的时间内，就是由 t 到 $t+\Delta t$ 一段时间内，将有什么变化．落体所已经经历的路程将得到某个增量 Δs．先前的路程 $s_1=\frac{gt^2}{2}$；增加后的路程

1) 公式(1)现在是从力学的一些普遍定律中推导出来的，但在历史上公式(1)恰恰是自实验确立并且是后来在这些定律中总结起来的经验的一部分．

$$s_2=\frac{g(t+\Delta t)^2}{2}=\frac{gt^2}{2}+\frac{g}{2}[2t\Delta t+(\Delta t)^2].$$

由此求得增量

$$\Delta s=s_2-s_1=\frac{g}{2}[2t\Delta t+(\Delta t)^2].$$

这就是由 t 到 $t+\Delta t$ 一段时间内所经历的路程. 以 Δt 除 Δs, 则得 Δs 一段路程上的平均速度:

$$v_{cp}=\frac{\Delta s}{\Delta t}=gt+\frac{g}{2}\Delta t.$$

令 Δt 趋近于零, 我们所得到的平均速度就将无限制地接近于 A 点上的真正速度. 另一方面, 我们看到上列等式的右方的第二项随着 Δt 的减小而变得乌有地小, 因此 v_{cp} 同时就趋于数量 gt. 这一结果通常是这样记写的:

$$v=\lim_{\Delta t\to 0}v_{cp}=\lim_{\Delta t\to 0}\frac{\Delta s}{\Delta t}=\lim_{\Delta t\to 0}\left(gt+\frac{g}{2}\Delta t\right)=gt.$$

所以 gt 就是在瞬时 t 的真正速度.

例 2 一个已经注满了水的贮水池有边长为 a 的正方形底和高度为 h 的垂直四壁(图 1). 问水对一壁所施的总压力有多大?

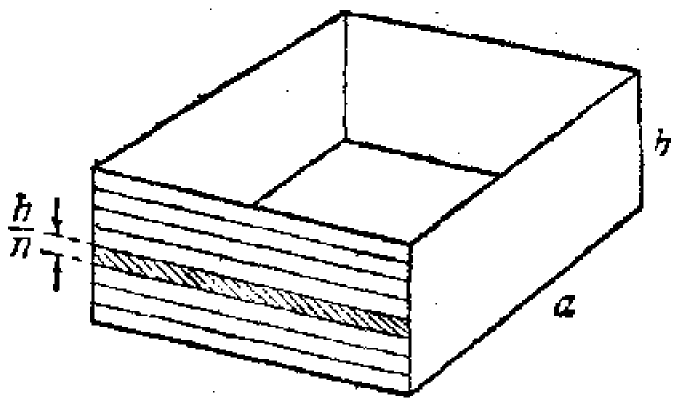


图 1

让我们把池壁的表面分成高度为 $\frac{h}{n}$ 的横条. 在容器每一点上的压力我们知道就跟位于这一点上面的液体柱的压力头相等. 所以在每一个横条下边的压力, 用适当的单位表出, 将分别等于 $\frac{h}{n}$, $\frac{2h}{n}$, $\frac{3h}{n}$, $\cdots$, $\frac{(n-1)h}{n}$, h. 若把每一个横条之内的压力当作固定不变, 我们就可以得到所求总力 P 的近似表达式. 由此可见 P 的近

似值等于

$$P\approx\frac{ah}{n}\cdot\frac{h}{n}+\frac{ah}{n}\cdot\frac{2h}{n}+\cdots+\frac{ah}{n}\cdot\frac{(n-1)h}{n}+\frac{ah}{n}h$$

$$=\frac{ah^2}{n^2}(1+2+\cdots+n)=\frac{ah^2}{n^2}\cdot\frac{n(n+1)}{2}$$

$$=\frac{ah^2}{2}\left(1+\frac{1}{n}\right).$$

要想求出力的真正数量，我们将把横条划分得愈来愈细，因而使 n 无限制地增大. 随着 n 的增大，在上式中数量 $\frac{1}{n}$ 将愈变愈小，而在极限中我们就得出准确的公式

$$P=\frac{ah^2}{2}.$$

极限法的观念是简单的，并可归结如下. 为了要确定某一个数量，我们最先加以确定的，不是这数量本身而是它的某些近似值. 这时，我们所建立的，也不只是一个近似值，而是一连串愈来愈准确的近似值. 然后从这一连串近似值的考察，也就是说，从这近似过程本身的考察，就把那数量的准确值唯一地确定下来了. 借助于这一种在实质上深刻的辩证的方法，我们认识到稳定不变的事物是过程、运动的结果.

数学极限法的创造是对于那些不能够用算术、代数及初等几何的简单方法来求解的问题进行了许多世代的顽强探索的结果.

究竟是在求解怎样的问题时才形成了数学分析的基本概念呢？对于这些问题的解决究竟创造了怎样的方法呢？

十七世纪的数学家已逐渐看出，把关于研究各种运动和量与量之间相倚性的大多数问题以及先前无法解决的几何问题归并起来，总不外乎两类. 第一类问题的最显著而简单的例子是：在非匀速运动中求某一瞬时的速度问题，与此相类似的关于量的变化率的问题，以及在曲线上求作切线的问题. 这些问题（我们的例 1 也在内）引导到数学分析中名为"微分学"的一部分. 第二类问题的最简单例子是：求曲线形的面积问题，在非匀速运动中求所经

历的总路程问题，以及一般地求连续变量的作用的总和问题(如上面所举的例 2). 这一群问题引导到数学解析的另一部分——积分学. 这样就突出了两个基本的问题：切线问题与面积问题.

在这一章中将要详细叙述那些观念给这两个问题的解决打下了基础. 特别重要的是牛顿与莱布尼茨定理，即是面积问题在某种意义上实是切线问题的反问题. 为了要解决切线问题，为了要解决可以归并于此的其他问题，已经找到一种便利而且十分普遍的计算法——显然可从而求到解答的普遍方法——导数法或微分法.

数学分析的建立与发展的历史以及笛卡儿所创造的解析几何学对于数学分析的萌芽所起的作用都已经在第一章里叙述过了. 我们看到在十七世纪的后半与十八世纪的前半整个数学发生了全面的变化. 在已有的数学各分支——算术、初等几何学、代数与三角的初阶——之外，增添了那样普遍的方法，如解析几何学、微分学以及附带有最简单的微分方程理论的积分学. 于是就有可能解决先前所不能梦想解决的问题了.

结果是：若曲线的规律不太复杂，则总可以在曲线的任何点上作出它的切线——只要借助于微分学的法则把所谓导数计算出来，而在许多场合这种计算只是几分钟的事. 在此之前，人们仅仅能够作圆的及另外少数曲线的切线并且完全不指望得到这问题的普遍解法.

若在任何给定时间内一点所经历的路程为已知，则用同样方法可以立即求出该点在任何给定瞬时的速度或加速度. 反之，应用微分法的倒转——就是所谓积分法——可以由加速度求出速度与路程. 据此，知道了椭圆的几何性质，就不难推证，由于牛顿运动定律和万有引力定律的结果，行星必须按照刻普勒定律沿着椭圆轨道绕日运动.

在实践中具有最重要意义的是关于最大与最小量的问题，也就是所谓极大与极小问题. 让我们引述这样一个例子：从半径为已知的圆材要砍出一根矩形截面的梁，使其有尽可能大的抗弯强

度．问梁的两边应该有怎样的比率？稍稍考虑一下矩形截面的梁的强度（应用到积分学的简单推理），并解出极大问题（为此须用到导数的计算），就可以获得这样的答案：当矩形截面的高与底成 $\sqrt{2}:1$ 的比率时强度为最大．极大与极小问题的解决正跟求作切线问题同样地简单．

在曲线的不同地点上，假使它既不是直线又不是圆线的话，曲线的弯曲程度，一般地说，是不相同的．怎样去计算跟已知曲线某一点附近弯曲得一样的那个圆的半径——也就是所谓曲线某一点上的曲率半径呢？这也同样地简单；只须两次应用微分法的运算就够了．曲率半径在力学的许多问题中起着重大的作用．

在新的演算法发明之前，人们只能够求出多边形、圆、圆扇形、圆弓形及其他少数图形的面积．此外，阿基米德还提供了计算抛物线弓形面积的方法[1]．他当时所使用的方法，以抛物线的特殊性质为根据，是富于机智的．由阿基米德的这一成就看来，我们不免要这样揣想：每一个新的面积计算问题就需要他的更机智而又更困难的探究．但是根据牛顿-莱布尼茨定理，切线问题的倒转可解决面积问题，因此也就立刻有可能把各种各样极不相同的曲线所围成的面积一一计算出来．当这一情况刚才显露端倪的时候，数学家的欢喜赞叹是不言而喻的．此其结果，就有了一种普遍的方法，适用于不可胜数的最不相同的图形．容积、曲面积、曲线长度、非均匀物体的重量等等的计算也是如此．

在力学中新方法的成就更为巨大．似乎没有一个力学上的问题是新的演算法所不能阐明，所不能解决的．

在此稍前的时候，巴斯卡已经把托利拆里真空在山上随着高度的上升而增大的现象解释为在上升时大气压力减小的结果．但压力的减小是按照怎样的规律的呢？这问题此时借助于简单微分方程的讨论便告解决了．

航海的人们都知道，只要把船缆在系船用的绞盘上缠绕一、二周，就已经一个人能够把一条大船靠码头带住．这是为什么呢？这

1） 关于这一方法，可参阅《数学通报》1957 年 1 月号第 14 页．——译者注

问题原来在数学上跟上一个问题几乎完全相同，因而可以立即得到解答.

所以在数学分析建立后，随着而来的是它的应用在技术与自然科学的最不相同的各种领域中蓬勃发展的时期. 经由特殊问题中某些特征的抽象化而创造出来的数学分析这样就反映着物质世界的十分现实和深刻的性质，而且正因为如此才成为对这样广大范围的种种问题进行研究的工具. 刚体的力学运动，流体和气体的运动，其个别粒子的运动和质量流动的规律，热与电的过程，化学反应的进程等等——所有这些现象都有相应的各门科学广泛地使用着数学分析的工具来加以研究.

在分析的应用领域扩大起来的同时，分析本身也无限量地丰富起来；分析的这些分支，如级数论、分析在几何上的应用、微分方程论，就一一产生并日趋完善.

在十八世纪中叶的数学家中间很流行着这样一种意见：自然科学上的任何问题，只要做得到从数学上来理解，也就是说，找得到它的正确的数学描述，就可以借助于解析几何学与微积分学而获得解决.

渐渐地开始出现了更复杂的需要把所说方法进一步发展的自然科学与技术问题. 为了要解决这些问题，就相继创立了数学的更多的分支：变分学、复变函数论、场论、积分方程、泛函分析，借以解决各类新的问题. 但是这些新的演算法在实质上都是十七世纪所发明的卓越演算法的直接延续和推广，十八世纪最伟大的数学家 D. 伯努利(1700—1782)、欧拉 (1707—1783) 与拉格朗日(1736—1813)，在开辟科学中的新途径时，就经常地从精确科学上的迫切问题出发. 数学分析的强有力发展在十九世纪也继续不已. 在那一世纪中生活并进行创作的，有那样著名的数学家，如高斯(1777—1855)、柯西(1789—1857)、奥斯特洛格拉特斯基(1801—1861)、车比雪夫(1821—1894)、黎曼(1826—1866)、阿贝尔(1802—1829)、魏尔斯特拉斯(1815—1897)，都在数学分析的发展中作出了极重大的贡献.

天才的俄罗斯数学家罗巴切夫斯基也影响了数学分析中某些问题的发展.

我们还要提一提生活在十九与二十世纪之际的最杰出数学家：马尔科夫(1856—1922)、李雅普诺夫(1857—1918)、庞加来(1854—1912)、克莱茵(1849—1925)、希尔伯特(1862—1943).

在十九世纪后半发生了对分析基础本身所进行的深刻而批判性的修正和更明确的定义. 已经积累起来的各种各样强有力的分析方法于是获得了共同的、有系统的、跟已经提高了的数学严密性水准相适应的基础. 正是所有这些方法, 连同算术、代数、几何与三角, 人们用来从数学上去理解周围的世界, 描述所发生的种种现象, 并且解决与这些现象有关的在实践上重要的问题.

在我们苏联, 自从伟大的十月社会主义革命以来, 数学分析, 跟整个的数学科学一起, 获得了特别充分而全面的发展和规模广泛的应用.

在所有的高等工业学校里都学习着解析几何、微积分及微分方程论, 因此数学的这些领域在苏联是千百万人所熟悉的; 在许多中等技术学校里则讲授着这些科学的初步知识; 在中等学校里怎样学习这些知识的问题也已经提了出来.

在最近的时期, 由于大型快速数学计算机的使用, 在数学中产生了一个新的转折点. 这种计算机, 跟所有的上述数学分支结合起来, 给人们提供了各种新的不平凡的可能性.

现在, 数学分析, 连同从分析中生长起来的各部分, 乃是一门枝叶繁茂的数学科学, 由彼此有紧密联系的广阔而独立的分支所构成, 这些分支个个在日趋完善并向前进展, 同时在这些成就中有相当大的部分是属于苏联的科学家的.

生活上的需要, 跟技术的宏大发展有关联的问题是数学分析的比先前任何时候更为巨大的主导力量. 其次轮到超音速航空动力学的数学问题, 这些问题的解决已逐渐地获得成就. 数学物理学上的一些最困难的问题现在已转入有可能求出实际数值解的阶段. 在近世物理学中, 像量子力学(而与此有关的就是微观世界现

象的理解问题)那样的理论不仅需要近世数学分析的最高深部分来解决它的问题，而且非用分析的工具无法说明它的基本概念.

这一章的目的在于用通俗的形式使只懂得初等数学的读者认识分析的那样基本、原始的概念，如函数、极限、导数与积分，是怎样发生的以及有哪些最简单的应用. 分析的一些特殊分支则将在本书的其他几章中加以表述. 因此这一章带有比较初等的性质，在已经熟悉寻常分析教程的读者尽可把它略去而无损于对后文的了解.

§2. 函　　数

函数概念　在自然界中各种物体和各种现象是有机地互相关联的，彼此相依赖的. 稳固不变的最简单关系，很早就有人加以研究了. 关于这种关系的知识逐渐积累起来，便形成为物理的定律. 在多数的场合，这指出了在数量上描述某些现象的几个不同的量是紧密地互相关联的，一个量完全决定于其他的量的值. 例如，矩形的两边的长短就把它的面积完全确定，已给的气体的体积在温度固定时决定于它的压力，已给的金属杆的伸长度决定于它的温度，等等. 类似的规律性就正是函数概念的根源.

代数公式使我们能够从进入公式的文字量的每一个值求出公式所表达的量的值，因而在这种公式中早已含有函数的概念了. 让我们来看几个由公式表出的函数的例子.

1. 设在起始的瞬时质点是静止的，然后由于重力的作用开始下坠. 于是质点在时间 t 内所经历的路程 s 就由下列公式表出：

$$s=\frac{gt^2}{2},\tag{1}$$

式中 g 是重力加速度.

2. 从边长为 a 的正方形做成了一个高为 x 的无盖小方盒(图 2). 这盒的容积 V 将按公式

$$V=x(a-2x)^2\tag{2}$$

来计算. 公式(2)使我们能够从每一个高度 x (这种 x 显然满足不

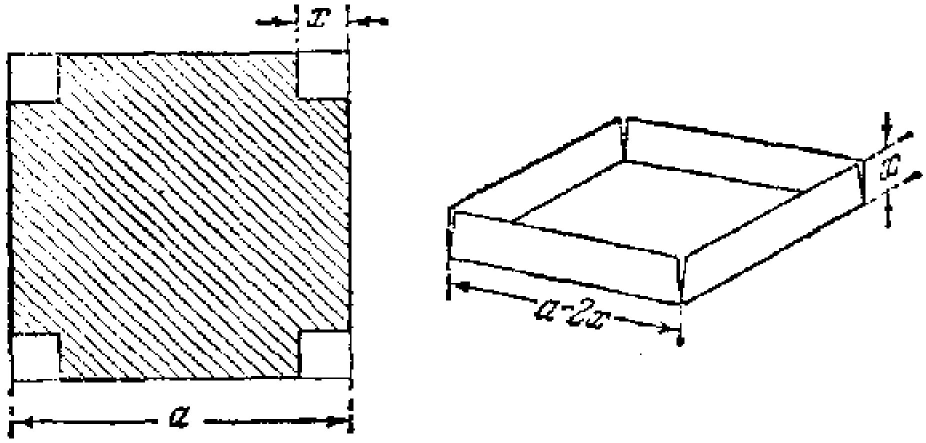


图 2

等式 $0\leqslant x\leqslant\frac{a}{2}$)求出盒的容积.

3. 设在圆形溜冰道的中心树立一柱,并在柱顶高达 h 的地方安上一灯(图3). 溜冰道的照度 T 就可以由下列公式表出:

$$T=\frac{A\sin\alpha}{h^2+r^2}, \tag{3}$$

式中 r 是溜冰道的半径, $\operatorname{tg}\alpha=\frac{h}{r}$, 而 A 是某一个表示灯光强度的量. 知道了高度 h, 我们可以从公式(3)把 T 算出.

4. 二次方程

$$x^2+px-1=0 \tag{4}$$

的根可由公式

$$x=-\frac{p}{2}\pm\sqrt{1+\frac{p^2}{4}} \tag{5}$$

来计算.

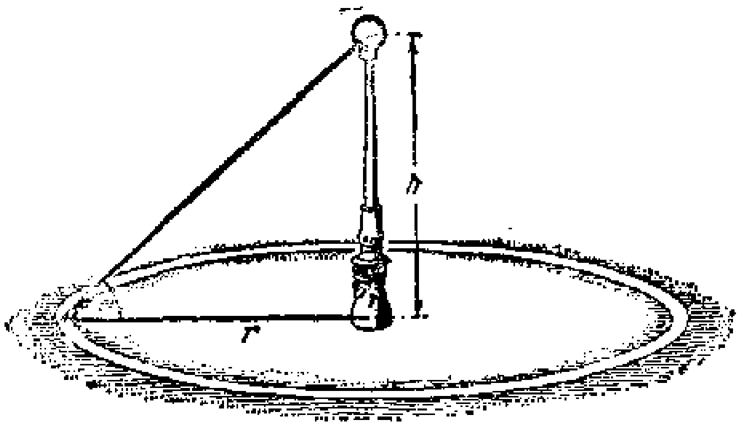


图 3

公式一般所具有的特征, 特别是上举各公式所具有的特征, 在于按照公式就有可能从一个量(时间 t, 盒高 x, 柱高 h, 方程的系数 p)所可

取的任何预先给定的值算出另一个量(路程 s, 容积 V, 照度 T, 方程的根 x)的相应值; 这里所说的第一个量称为自变量, 而第二个量称为因变量或第一个量的函数.

所举各公式都提供了我们函数的例子: 质点所经历的路程 s 是时间 t 的函数; 盒的容积 V 是盒高 x 的函数; 溜冰道的照度 T 是柱高 h 的函数; 二次方程(4)的两个根是系数 p 的函数.

必须指出, 在有一些场合, 自变量可以取任何预先给定的数值, 如例 4 中的自变量, 即二次方程 (4) 的系数 p, 便可以是任意的数. 在另外一些场合, 自变量只从某一个预先确定的数集中取任何一值, 如在例 2 中盒的容积是盒高 x 的函数, 其自变量可以从满足不等式 $0\leqslant x\leqslant \frac{a}{2}$ 的数集中取任何一值. 正好同样地, 在例 3 中溜冰道的照度 T 是柱高 h 的函数, 其自变量在理论上可取满足不等式 $h>0$ 的任何一值, 但在实际上只能取满足不等式 $0<h\leqslant H$ 的任何一值, 这里的量 H 取决于溜冰场管理处所掌握的技术可能性.

让我们再举一些这样的例子: 公式

$$y=\sqrt{1-x^2}$$

确定了一个实函数, 它所表达的是实数 x 与 y 之间的对应关系, 但显然这不是对所有的 x, 而只是对那些满足不等式 $-1\leqslant x\leqslant 1$ 的 x 而言. 至于公式

$$y=\lg(1-x^2)$$

所确定的实函数则以满足不等式 $-1<x<1$ 的 x 值为限.

因此, 我们须注意到这一情况: 具体的函数不一定可以对 x 的所有可能的数值给出, 而只可以限于 x 的数值的某一个集, 往往就是 x 轴上的某一段(包括或不包括端点在内)[1].

现在我们可以把目前在数学中所通用的关于函数概念的定义提出来了.

若有一个法则存在, 使得属于某一数集的每一个 x 值有一个

1) 这样的线段称为区间, 包括端点在内的, 如 $a\leqslant x\leqslant b$, 称为闭区间, 往往记作 $[a, b]$; 不包括端点的, 如 $a<x<b$, 称为开区间, 记作 (a, b). ——译者注

确定的 y 值相对应，则量 y 就是(自变)量 x 的一个(单值)函数.

在这定义中居于显著地位的 x 值的集称为函数的定义域.

每一种新的概念常常引起新的符号，从算术转变到代数的关键在于有可能立出对任何数据适用的公式，——这种普遍解的寻求便导致了文字符号.

数学分析的课题在于研究种种函数——一个量与另一个量的种种相倚性；正像在代数中从具体的数转变到任意的数——文字，在分析中我们从具体的公式转变到任意的函数. "y 是 x 的函数"一语，我们将依照惯例写成：

$$y=f(x).$$

正像在代数中对于不同的数使用不同的文字，在分析中对于不同的相倚性——函数——也使用不同的记法，如 $y=F(x)$, $y=\varphi(x)$, ….

函数的图形 十七世纪后半最有成效和辉煌的观念之一就是函数概念与几何线条之间有着联系的观念. 要实现这种联系，比如说，可以使用笛卡儿直角坐标系，这自然是读者早已从中学数学教程里粗略地有所认识的.

让我们在平面中给定一个笛卡儿直角坐标系. 这就是说，我们在这平面中选定两条互相垂直的直线(横轴与纵轴)，并且在每一条直线上把正方向确定，于是对应于平面的每一点 M 可以找到两个数 (x, y)——它的坐标，分别以已选定的尺度表达了点 M 到纵轴与横轴的距离，但带有相应的正负号[1].

借助于这种坐标系我们就可以用图形把函数表示为某些线条了. 设已给定某一函数

$$y=f(x). \tag{6}$$

我们知道这就意味着，对于函数定义域中每一个已给的 x，可以用某种方法把相应的 y 值确定，比如说，直接算出. 我们将使 x 取所有可能的数值. 对每一个 x，让我们按照法则(6)把 y 确定，并在平面中作出具有坐标 x 与 y 的点. 这样，在 x 轴的每一点 M'

1) 数 x 与 y 分别称为 M 的横坐标与纵坐标.

之上[1]，有着坐标为 x 与 $y=f(x)$ 的点 M（图 4）。所有的点 M 的集合就构成某一线条，我们将称之为函数 $y=f(x)$ 的图形。

因此，所谓函数 $f(x)$ 的图形，就是坐标满足方程(6)的点的几何轨迹。

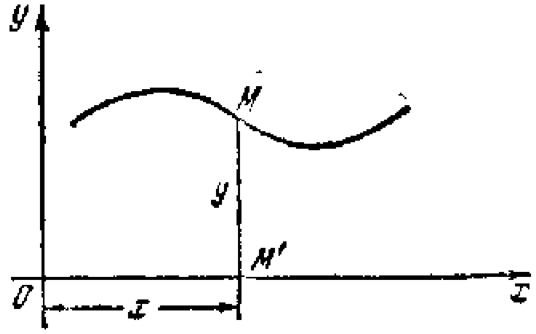


图 4

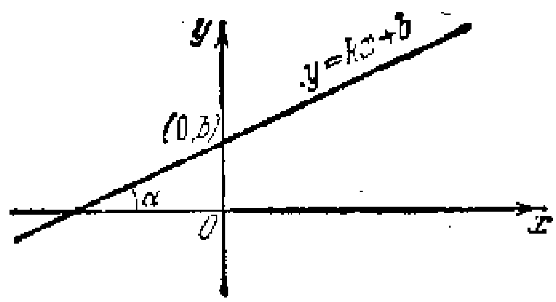


图 5

在中学里我们已经认识了最简单函数的图形。例如，读者想必知道函数 $y=kx+b$，其中 k 与 b 是常数，用图形表出，就是一条直线(图 5)，它与 x 轴的正方向构成一个满足 $\operatorname{tg}\alpha=k$ 的夹角 α，并与 y 轴相交于点 $(0,\ b)$。这个函数有线性函数之称。

线性函数在应用上可特别频繁地碰到。我们都记得有许多物理定律相当准确地由线性函数表出。例如，物体长度 l 可以很近似地当作是物体温度的线性函数：

$$l=l_0+\alpha l_0 t,$$

式中 α 是线膨胀系数，l_0 是物体在 $t=0$ 时的长度。若 x 是时间而 y 是一点在时间 x 内所经历的路程，则线性函数 $y=kx+b$ 显然就表示该点正在以速度 k 作匀速运动；而数 b 所记的是在 $x_0=0$ 的瞬时我们的点与路程起算处的距离。线性函数所以得到极广泛的使用，不仅是由于这函数本身的简单，也因为不均匀的变化至少在小的分段上有近似地当作均匀的可能。

在其他的场合则必须应用另外的函数相倚性。例如，我们都记得玻义耳-马里奥特定律：

---

[1] 这里显然假定 y 得正值。若 y 值为负，则 M 将在 M' 之下。——译者注

$$v=\frac{c}{p},$$

式中 p 与 v 之间的相倚性乃是这两个量的一种反比关系，这种相倚性的图形就是双曲线(图 6)．

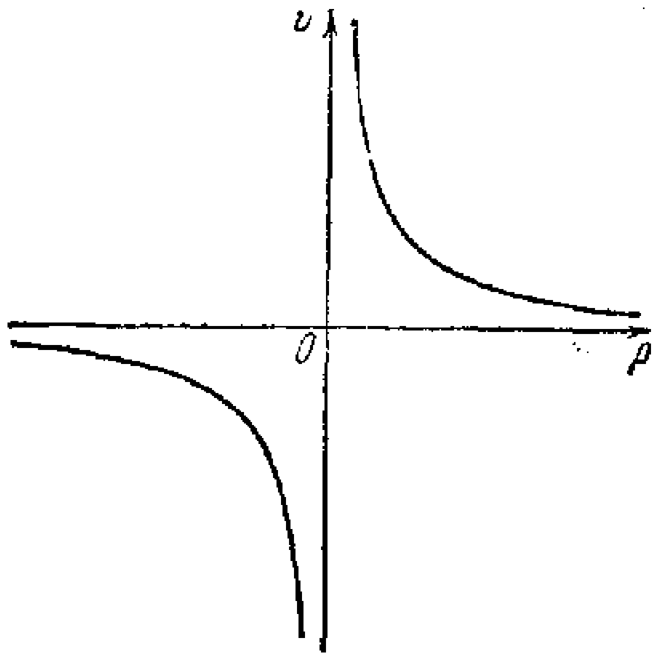


图 6

单就玻义耳-马里奥特物理定律而论，这相当于 p 与 v 都取正值的场合，在图形中就由双曲线在第一象限内的分支来描述.

振动过程的场合是伴有周期运动的，而周期运动则常常由我们已知是周期地变化的三角函数来描述．例如，把一个悬挂着的弹簧在其弹性限度内拉长，因而使其失去平衡，那末弹簧的一点 A 将作垂直的振动，可由下列定律颇准确地表出：

$$x=a\cos(pt+\alpha),$$

式中 x 是 A 点跟平衡位置的距离，t 是时间，而数 a, p 及 α 是某些常量，分别由弹簧的物质、大小及初应力来确定.

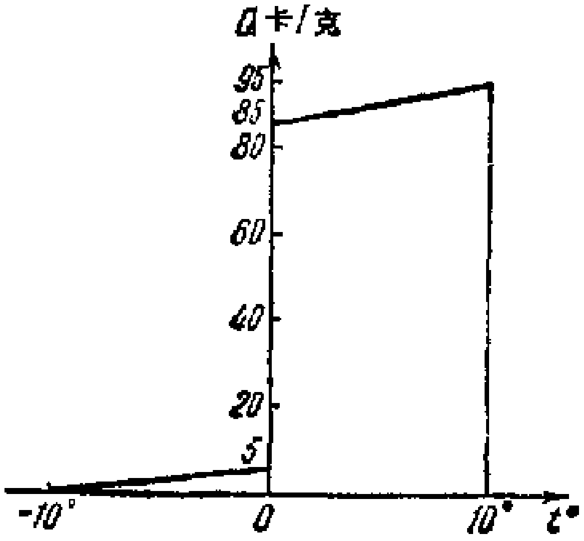


图 7

应当注意到一个函数可以在不同的区域中用不同的公式来定义，而且这是可以取决于事物的具体情况的．例如，一克水(冰)的温度 t 与其中所含的热量 Q 之间的相倚性 $Q=f(t)$，当 t

变动于 -10° 与 $+10^\circ$ 之间时，是一个完全确定而难以用单独一个公式表出的函数[1]，但这函数是易于由两个公式给出的. 由于冰的热容量等于 0.5 而水的热容量等于 1，若在一定条件下取 -10° 时的量 $Q=0$，则这函数就可以在 t 变动于区间 $-10^\circ \leqslant t < 0^\circ$ 时由公式

$$Q=0.5t+5$$

表出，在 t 变动于区间 $0^\circ \leqslant t \leqslant 10^\circ$ 时由另一公式

$$Q=t+85$$

表出. 当 $t=0$ 时，这函数是不定的——多值的；为便利起见可以假定当 $t=0$ 时它取一个完全确定的值，比如说，$f(0)=45$. 函数 $Q=f(t)$ 的图形已在图 7 中绘出.

我们已经举了许多由公式给出的函数的例子. 借助于公式来给出函数，从数学的观点看来，是最重要的；因为当函数由公式给出时，对于用数学方法来讨论函数的性质，具备着最有利的条件.

但不要以为公式是给出函数的唯一方法. 我们还有许多其他的方法，其中具有特殊意义的是函数的图形，它提供了函数的一目了然的几何表现. 这可以从下面的例子得到很好的说明.

为了要知道空气的温度在一昼夜间怎样变化，在气象台上使用着一种仪器，叫做温度记录器. 这种记录器的主要部分是借助于时钟机构而绕轴自转的圆筒和对温度变化极为灵敏的弓形黄铜匣. 当温度增高时铜匣自行伸直，因此借杠杆装置而连结在匣上的自动书写笔就会上升. 反之，温度的减低则使笔下移. 在圆筒上适当地缠着划有格线的纸带，而自动笔就在这纸带上绘出连续不断的线条——函数 $T=f(t)$ 的图形，表达了时间与空气温度之间的相倚性. 借助于所得到的图形，用不着计算，便可以确定每一瞬时 t 的温度 T 了.

刚才所举的例子表明图形可以单独地确定函数，不管它是不

1) 这并不是说，这样的表达不可能. 在第二卷第十二章中将指出怎样求得单独的公式.

是由公式给出.

可是，我们将来还要回到这个问题(见第二卷第十二章)并证明下面这个极重要论断的正确性，每一个连续的图形可以用某一个公式，或者照仍然通行的说法，某一个分析式来表达，对于很多不连续的图形说来，这也是确实的[1].

我们要指出，这个具有重大原则性意义的论断还只在上一世纪的中叶才在数学里被充分领会．在此时之前，数学家们把"函数"一词仅仅理解为分析式(公式)，但同时他们错误地以为决不是每一个连续的图形有相应的分析式，并且推想，若一个函数由公式给出，则它的图形，跟其他的图形相比，应当具有特别美好的性质.

但在十九世纪发现了所有的连续图形都可以由或多或少地复杂的公式给出．因此分析式在作为确定函数的一种方法时所独有的作用已告动摇，结果便形成了新的、更灵活的、关于函数概念的定义，就是我们在上文所提出的定义．根据这个定义，要把变量 y 称为变量 x 的函数，只须有一个法则存在，使得在这函数的定义域中的每一个 x 值有一个完全确定的 y 值相对应就是，不管这一法则由怎样的方法给出：公式、图形、表格或还有任何其他的方法.

在这里不妨指出，上述定义在数学文献中往往跟数学家狄利克雷的名字连在一起．应当着重说明，跟狄利克雷同时，这个定义也曾由罗巴切夫斯基独立地提出.

最后，我们建议读者绘出下列各函数的图形作为一种练习：x^3，$\sqrt{x}$，$\sin x$，$\sin 2x$，$\sin\left(x+\frac{\pi}{4}\right)$，$\ln x$，$\ln(1+x)$，$|x-3|$，$\frac{x+|x|}{2}$.

读者也应当看得很清楚，使关系式

$$f(-x)=f(x)$$

对一切 x 值成立的函数有对称于 y 轴的图形；而在关系式

1) 自然，只有当"公式"与"分析式"两个名词在数学中究竟作何解释被确切地规定以后，读者才会完全明了上面所说的论断.

$$f(-x)=-f(x)$$

的场合，则其图形对称于坐标原点. 试想一想怎样可以从 $f(x)$ 的图形得到 $f(a+x)$ 的图形，其中 a 是一个常数. 最后，试探究怎样可以借助于 $f(x)$ 与 $\varphi(x)$ 的图形找到复合函数 $y=f[\varphi(x)]$ 的值.

§3. 极　　限

在§1里已经说过，近世数学分析运用着一种独特的方法，这方法经历了许多世纪方才炼成而它正是分析中的基本推论工具. 这里所说的就是无穷小法，或者照实质上完全相同的说法，就是极限法. 我们将力求说明这些概念. 为此，让我们来看一看下面的例子.

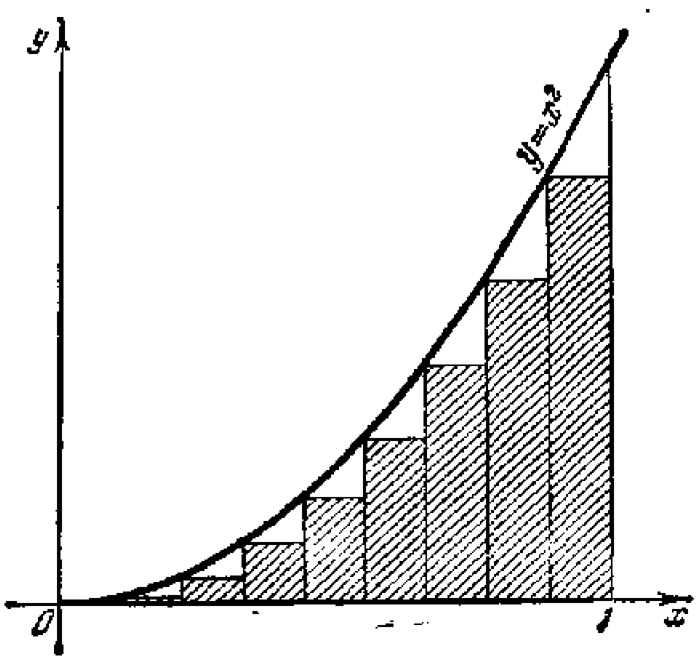


图 8

我们要计算由抛物线 $y=x^2$，x 轴以及直线 $x=1$ 所围成的面积(图 8). 初等数学没有提供我们解决这一问题的方法. 但这正是我们可以在这里做到的.

让我们用下列各点：

$$0,\ \frac{1}{n},\ \frac{2}{n},\ \cdots,\ \frac{n-1}{n},\ 1$$

把 x 轴的一段[0, 1]分成 n 个相等的小段，并在每一个小段上作出左上角碰到抛物线的矩形. 结果我们就得到图 8 中加有阴影的一连串矩形，其面积的总和 S_n 等于：

$$S_n = 0\,\frac{1}{n} + \left(\frac{1}{n}\right)^2 \frac{1}{n} + \left(\frac{2}{n}\right)^2 \frac{1}{n} + \cdots + \left(\frac{n-1}{n}\right)^2 \frac{1}{n}$$
$$= \frac{1^2+2^2+\cdots+(n-1)^2}{n^3} = \frac{(n-1)n(2n-1)}{6n^3}\text{[1]}.$$

让我们把量 S_n 写成下面的样子：

$$S_n = \frac{1}{3} + \left(\frac{1}{6n^2} - \frac{1}{2n}\right) = \frac{1}{3} + \alpha_n. \tag{7}$$

随着 n 而定的量 α_n，虽然形式颇繁，却具有一种值得注意的性质：若 n 无限制地增大，则 α_n 将趋于零．这一性质还可以这样来表述：给定任意的正数 ε，便可挑出这样大的值 N，使得就大于 N 的一切 n 而论，数 α_n 的绝对值将小于已给的 ε[2]．

就近世数学中所理解的无穷小量而言，量 α_n 就是无穷小量的一个例子．

从图 8 我们看到，若数 n 无限制地增大，则加有阴影的矩形面积的总和 S_n 将趋于所求曲线形的面积．另一方面，由于 α_n 在 n 无限制增大时趋于零，等式(7)就显示总和 S_n 同时趋于 $\frac{1}{3}$．由此可知所求的曲线形面积等于 $\frac{1}{3}$，而我们的问题已告解决了．

所阐述的方法因此可归结如下．为了要寻求某一个量 S，我们引用了另外一个向它趋近的变量 S_n，这变量按照某一种法则

1) 在显而易见的等式 $(k+1)^3-k^3=3k^2+3k+1$ 中，依次令 $k=1, 2, \cdots, n-1$，然后把各等式的左方与右方分别相加并令 $\sigma_n=1^2+2^2+\cdots+(n-1)^2$，可得 σ_n 的方程 $n^3-1=3\sigma_n+\frac{3(n-1)n}{2}+n-1$，把这方程解出，得

$$\sigma_n = \frac{(n-1)n(2n-1)}{6}.$$

2) 例如，设 $\varepsilon=0.001$，则可取 $N=500$．事实上，因为

$$\frac{1}{6n^2} < \frac{1}{2n}$$

对于整数的 n 成立，所以当 $n>500$ 时有

$$|\alpha_n| = \left|\frac{1}{6n^2} - \frac{1}{2n}\right| = \frac{1}{2n} - \frac{1}{6n^2} < \frac{1}{2n} < 0.001.$$

类似地可以给定任意小的 ε 值，比如说：

$$\varepsilon_1 = 0.0001,\ \varepsilon_2 = 0.00001,\ \cdots,$$

并且像上面一样选择与此相应的值 $N=N_1, N_2, \cdots$．

随着自然数 $n=1, 2, 3, \cdots$ 而经历个别的特殊值 $S_1, S_2, S_3, \cdots$，于是，在指出了变量 S_n 可写成常数 $\frac{1}{3}$ 与无穷小量 α_n 的和式以后，我们就断定 S_n 趋于 $\frac{1}{3}$，因而 $S=\frac{1}{3}$. 使用近世极限论的语言于这一场合，我们可以说，当 n 无限制增大时，变量 S_n 趋向一个极限，而这极限等于 $\frac{1}{3}$.

现在让我们给所说概念提出一个准确的定义.

若变量 $\alpha_n (n=1, 2, \cdots)$ 具有这样的性质：对于无论怎样小的一个正数 ε，总可以挑出那样一个足够大的 N，使得不等式 $|\alpha_n|<\varepsilon$ 在 $n>N$ 时成立，则我们便说，α_n 是一个**无穷小量**[1]，并记作

$$\lim_{n\to\infty} \alpha_n = 0 \quad 或 \quad \alpha_n \to 0.$$

另一方面，若某一变量 x_n 可以写成和式，

$$x_n = a + \alpha_n,$$

其中 a 是某一个常数而 α_n 是无穷小，则我们便说，变量 x_n 在 n 无限制增大时趋于数 a，并记作

$$\lim x_n = a \quad 或 \quad x_n \to a.$$

数 a 称为 x_n 的**极限**. 特别是无穷小量的极限显然是零.

让我们来看一些变量的例子：

$$x_n = \frac{1}{n}, \quad y_n = -\frac{1}{n^2}, \quad z_n = \frac{(-1)^n}{n}, \quad u_n = \frac{n-1}{n} = 1 - \frac{1}{n};$$

$$v_n = (-1)^n \quad (n=1, 2, 3, \cdots).$$

显然，x_n, y_n 与 z_n 都是无穷小，其中第一个渐减地趋于零，第二个始终取负值，渐增地趋于零，而第三个振荡于零的左右而趋于零. 其次，$u_n \to 1$ 而 v_n 根本没有极限，因为随着 n 的增加并不趋近任何常数，却始终振荡，时而取值 1，时而取值 -1.

在分析中起着重大作用的还有**无穷大量**[2] 的概念. 所谓无穷

---

1) 无穷小量往往简称为无穷小.——译者注

2) 无穷大量也往往简称为无穷大.——译者注

大量，我们定义为具有下述性质的变量 $x_n(n=1, 2, 3, \cdots)$，对于无论怎样大的一个正数 M，总可以说出那样一个 N，使得

$$|x_n|>M$$

在 $n>N$ 时成立.

量 x_n 是无穷大这一事实我们记作

$$\lim x_n=\infty \text{ 或 } x_n\to\infty.$$

关于这样的量 x_n，我们说它趋于无穷大. 若同时从某一个值起这个量就总是正的(或负的)，则记作 $x_n\to+\infty$(或 $x_n\to-\infty$). 例如，当 $n=1, 2, 3, \cdots$,

$$\lim n^2=+\infty, \lim(-n^2)=-\infty;$$

$$\lim \lg\frac{1}{n}=-\infty, \lim \operatorname{tg}\left(\frac{\pi}{2}+\frac{1}{n}\right)=-\infty.$$

不难看出，若量 α_n 是无穷大，则 $\beta_n=\frac{1}{\alpha_n}$ 是无穷小，而反之亦然.

两个变量 x_n 与 y_n 可以彼此相加、相减、相乘及相除，并且，一般地说，就得出新的变量：和 x_n+y_n，差 x_n-y_n，积 x_ny_n 及商 $\frac{x_n}{y_n}$. 它们将分别取得下列特殊值：

$$x_1\pm y_1, x_2\pm y_2, x_3\pm y_3, \cdots$$

$$x_1y_1, x_2y_2, x_3y_3, \cdots$$

$$\frac{x_1}{y_1}, \frac{x_2}{y_2}, \frac{x_3}{y_3}, \cdots.$$

读者也可以独自证明这样一个颇为明显的事实：若变量 x_n 与 y_n 趋于有尽的极限，则这两个变量的和、差、积、商也各趋极限，而且所趋极限就分别等于原来两个极限的和、差、积、商. 这可以记写如下：

$$\lim(x_n\pm y_n)=\lim x_n\pm\lim y_n;$$

$$\lim(x_ny_n)=\lim x_n \lim y_n;$$

$$\lim \frac{x_n}{y_n}=\frac{\lim x_n}{\lim y_n}.$$

只有在商的场合，我们必须假定分母的极限($\lim y_n$)不等于零．若 $\lim y_n=0$ 而 $\lim x_n\neq 0$，则 x_n 与 y_n 的比已经不趋于有尽的极限而趋于无穷大了．

若分子与分母同时都趋于零，则将发生一种极有趣而又重要的场合．在这种场合，无法预先说出，比 $\frac{x_n}{y_n}$ 是否将趋于一个极限，因为这问题的答案完全取决于 x_n 与 y_n 向零趋近的特征．例如，若

$$x_n=\frac{1}{n},\quad y_n=\frac{1}{n^2},\quad z_n=\frac{(-1)^n}{n}\quad (n=1,\ 2,\ 3,\ \cdots),$$

则
$$\frac{y_n}{x_n}=\frac{1}{n}\to 0,\quad \frac{x_n}{y_n}=n\to\infty.$$

另一方面，量

$$\frac{x_n}{z_n}=(-1)^n$$

显然不趋于任何极限．

这样，比式的分子与分母都趋于零的场合是不可能用一般的定理来预先说明的；对于每一个这样的比式必须各别地进行特殊的探讨．

以后我们将看到，微分学的基本问题，也可以说就是在非匀速运动中求某一瞬时的速度问题，归结于确定两个无穷小量——路程增量与时间增量——之比的极限．

以上我们考虑了这样的一种变量 x_n，它在下标 n 无限制地经历一系列的自然数 $n=1,\ 2,\ 3,\ \cdots$ 时取得一连串的数值 $x_1,\ x_2,\ x_3,\ \cdots,\ x_n,\ \cdots$．但是我们也可以考虑 n 连续地变化，像时间一样，并且在这一条件下类似地来定义变量 x_n 的极限．这种极限的性质是跟上面对离散(即不连续)变量所陈述的性质完全相类似的．还要注意 n 无限制增大的情况在这里并不怎样重要，我们可以同样有效地考虑 n 在连续地变化时趋近于某一个已给定的值 n_0．

作为一个例子，让我们来考察量 $\frac{\sin x}{x}$ 在 x 趋近于零时的变化．左表是当 x 取某些值时这一个量的数值（x 的值假定是用弧度表出的）．

x	$\frac{\sin x}{x}$
0.50	0.9589…
0.10	0.9983…
0.05	0.9996…
…	…

显然，随着 x 的趋近于零，量 $\frac{\sin x}{x}$ 趋于1，但是这当然还需要严格的证明．要得到这个证明，比如说，可以从不等式

$$\sin x < x < \operatorname{tg} x$$

出发，这不等式是对于第一象限内所有异于零的角都成立的．用 $\sin x$ 除不等式的所有部分，得

$$1 < \frac{x}{\sin x} < \frac{1}{\cos x},$$

因此，

$$\cos x < \frac{\sin x}{x} < 1.$$

但是因为 $\cos x$ 随着 x 的向零减小而趋于1，所以包含在 $\cos x$ 与1之间的量 $\frac{\sin x}{x}$ 也趋于1，即

$$\lim_{x\to 0} \frac{\sin x}{x} = 1.$$

我们在后面就有机会要应用到这一个情况．

上面的等式我们是就 x 始终取正值而趋于零的场合来证明的．对这证明略加显而易见的改变，也便可以适用于 x 取负值而趋于零的场合．

现在我们还要讨论一个问题．变量可以有极限，也可以没有极限．因此产生这样一个问题：我们能不能给出一些准则，借以断定哪些变量确有极限存在．在可能给出准则的各种场合中，有一种重要而且足够普遍的，我们要在这里说一说．试想像变量 x_n 是增大的，或者至少是不减小的，也就是说满足不等式

$$x_1 \leqslant x_2 \leqslant x_3 \leqslant \cdots$$

的，并且还发现变量的所有这些值不超过同一个数 M，即 $x_n \leqslant M$

$(n=1, 2, 3, \cdots)$. 若把 x_n 诸值及数 M 都在 x 轴上标出，则可见动点 x_n 在轴上向右移动而始终在点 M 的左面. 很显然，变点 x_n 必然要趋于某一个极限点 a，这一点或者位于 M 点的左面，或者在极端的场合跟 M 相合.

这样，在所考虑的场合，我们的变量有极限

$$\lim x_n = a$$

存在.

上述推论带有直观性，但是不能够当作证明. 在现代的数学分析教程中，以实数理论作为基础，都提供了这一事实的完备的根据.

作为一个例子，让我们来考虑变量

$$u_n=\left(1+\frac{1}{n}\right)^n \quad (n=1, 2, 3, \cdots).$$

它的最初几个值 $u_1=2$, $u_2=2.25$, $u_3\approx 2.37$, $u_4\approx 2.44$, $\cdots$, 我们看得出是增大的. 把我们的式子按牛顿二项式展开后，可以证明它对 n 的任何值都是增大的. 而且还可以轻而易举地证明，不等式 $u_n<3$ 对一切 n 成立. 在这样的场合，我们的变量必然有不超过数 3 的极限. 我们将在后面看到，这个极限在数学分析中起着很重大的作用，在某种意义上说来正是对数的最自然的底.

这个极限通常用字母 e 记出. 它等于

$$e=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n=2.718281828459045\cdots.$$

更详细的研究证明了数 e 不是有理数[1].

还可以证明，所考虑的极限不仅在 $n\to\infty$ 时而且在 $n\to-\infty$ 时存在并等于 e. 同时，在两种场合，n 都可以不仅仅经历整数的值.

我们还要说一说极限概念在自然科学中的一种重要作用. 这作用就在于这样一件显著的事实：只有借助于极限概念（经过取极

---

1) 关于这一点，我们应当指出，有理数，即形如 $\frac{p}{q}$ 的数，其中 p 与 q 都是整数，经过加、减、乘、除（但不能以零作除数）的运算后，结果还是有理数. 但在极限运算中这就不一样了. 有理数序列的极限可以是无理数.

限的步骤)我们才得以对自然科学中所碰到的许多具体的量给出完全而详尽的定义.

姑且让我们来考虑下面这个几何的例子. 在中学的几何教程里我们先研习由直线段所围成的图形. 然后提出比较困难的关于寻求有已知半径的圆周长度的问题.

把求解这问题时所存在的困难加以分析,可归结如下.

我们必须看得很清楚,什么是圆周长度,也就是说,我们必须给出它的准确定义. 要紧的是这定义应当归结到直线段的长度,并且从这定义也应当有可能把圆周长度有效地算出.

不言而喻,计算的结果应该跟实践吻合. 比如说,我们有一个由实物(例如纱线)所做成的圆周. 把这圆周剪断、拉直,我们应该得到一条直线,它的长度在丈量的准确限度内应该跟计算相合.

从中学教程里我们已经知道这问题的解决归结为下面的定义. 所谓圆周长度,就是圆内接正[1]多边形的周长当边数无限制增大时的极限. 由此可见,所提出的问题在实质上是根据极限概念来解决的.

任意光滑曲线的长度正是类似地加以定义的. 在以下的几节里我们将看到几何与物理量的许多例子,这些量的准确定义也幸亏应用了极限概念方才能够给出.

极限与无穷小的概念是在上一世纪的初叶最后形成的. 它的定义,像上面所提出的那样,是跟柯西的名字连在一起的. 在柯西之前,在数学中还运用着比较不明确的概念. 现代的关于极限的、关于作为变量看待的无穷小的、关于实数的概念都是数学分析随着它的成就的确立和巩固而逐渐发展的结果.

§4. 连续函数

在数学分析所运用的函数中,连续函数构成了主要的一类. 要

1) 多边形的正不正无关紧要. 在这里,重要的只是圆内接可变多边形的最大边应当趋于零.

得到连续函数的概念，我们不妨这样说，它的图形是连续的，也就是可以一笔画成的.

在数学上连续函数表现出这样一种我们在实践中常常会碰到的性质，就是跟自变量的微小增量相对应的是因变量(函数)的微小增量. 物体运动的各种规律 $s=f(t)$，表达了物体所经历的路程 s 与时间 t 的相倚性，就可以作为连续函数的绝妙例子. 时间与空间都是连续的，同时物体运动的这种或那种规律 $s=f(t)$ 在两者之间确立了一定的连续的联系，其特征就在于跟时间的微小增量相对应的是路程的微小增量.

人们是在观察四周围所谓连绵不绝的介质——固体、液体或气体的，例如金属、水、空气——的时候达到连续性的抽象概念的. 事实上，如现在所周知，每一种物理的介质是由数目众多的相对运动着的单个粒子积累而成. 但这些粒子以及它们之间的距离，跟在宏观的物理现象中所会碰到的介质的体积相比较，是那样的微小，以致许多这样的现象可以充分完善地加以研究，只要把所研究的介质的质量当作是没有任何空隙而连续地分布在它所占据的空间就够了. 许多物理学的分支，例如流体动力学、气体动力学、弹性理论，就是以这样的假设作为基础的. 数学的连续性概念自然在这些以及其他许多分支中起着重大的作用.

让我们来考虑任何一个函数 $y=f(x)$ 以及自变量的一个完全确定的值 x_0. 若我们的函数反映着某一种连续过程的话，则对应于跟 x_0 相差很小的 x 值应该是跟函数在 x_0 一点上的值 $f(x_0)$ 相差很小的函数值 $f(x)$. 因此，若自变量的增量 $x-x_0$ 很小，则相应的函数增量 $f(x)-f(x_0)$ 也应该很小. 换句话说，若自变量的增量 $x-x_0$ 趋于零，则函数增量 $f(x)-f(x_0)$ 也应该趋于零，并可记写如下：

$$\lim_{x-x_0\to 0}[f(x)-f(x_0)]=0. \tag{8}$$

这一关系式就是函数在 x_0 一点上连续的数学定义.

函数 $f(x)$，若满足等式(8)，就称为**在 x_0 一点上连续**.

我们还要提出这样一个定义：

一个函数，若在已给区间的每一点 x_0 上连续，也就是说，若在每一个这样的点上满足等式 (8)，就称为对这区间的所有值连续.

这样，为了要给函数的这样一种性质，即它的图形是一条连续的（按照这一用语的寻常解释）曲线，提出数学的定义，我们必须先给局部的连续性（在 x_0 一点上的连续性）下定义，然后在这基础上来定义函数在整个线段上的连续性.

上述定义，最先由柯西于上世纪初指出，已在现代的数学分析中被普遍采用. 通过无数的具体例子的验证，显示了这一个定义恰好符合于我们所原已形成了的关于连续函数的实际概念，比如说，关于连续图形的概念.

读者从中学数学里所早已熟悉的初等函数 x^n, $\sin x$, $\cos x$, a^x, $\lg x$, $\arcsin x$, $\arccos x$ 都可以作为连续函数的例子. 所有这些函数都是在它们的定义域上连续的.

若把连续函数相加、相减、相乘、相除（在分母不等于零时），则结果又是连续函数. 但在相除时，就那些使分母中的函数变为零的值 x_0 而论，连续性一般地都遭到破坏. 所以相除的结果是一个在 x_0 点上间断的函数.

函数 $y=\frac{1}{x}$ 在 $x=0$ 一点上间断，就是这种函数的一个例子. 图 9 中所绘的图形是间断函数的另外一些例子.

我们建议读者仔细地把这些图形看一看. 我们要指出，函数的间断点是各种各样的：有时候当 x 趋近于函数不连续的 x_0 点时，极限 $f(x)$ 虽存在而异于 $f(x_0)$；有时候，如在图 9c 中，这极限就根本不存在. 也有这样的情况，当 x 从一面趋近 x_0 时，$f(x)-f(x_0)\to 0$，而当 x 从另一面趋近 x_0 时，则 $f(x)-f(x_0)$ 已不趋于零. 在这种场合，自然我们还是有函数的间断点，虽则对于这个函数我们可以说它是在这一点上“一面连续的”. 所有这些场合都可以在所说的图形中看到.

作为一种练习，我们建议由读者自己来回答下面的问题：函数

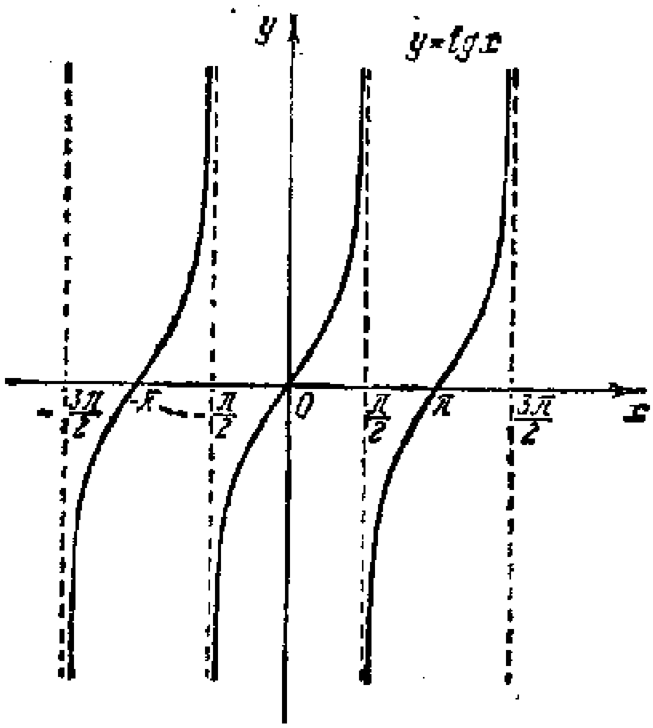

y
y=tg x
x
$-\frac{3\pi}{2}$
$-\pi$
$-\frac{\pi}{2}$
0
$\frac{\pi}{2}$
π
$\frac{3\pi}{2}$

图9 a

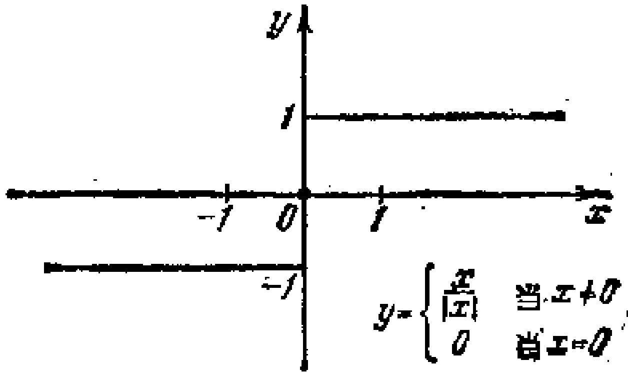

y
1
-1
0
1
x
-1
$y=\begin{cases}\frac{x}{|x|} & 当 x\neq 0\\ 0 & 当 x=0\end{cases}$

图9 b

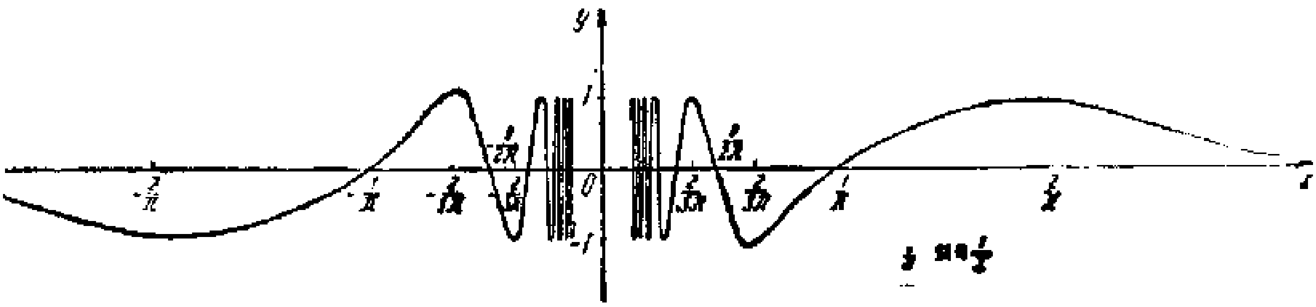

y
1
-1
0
x

图9 c

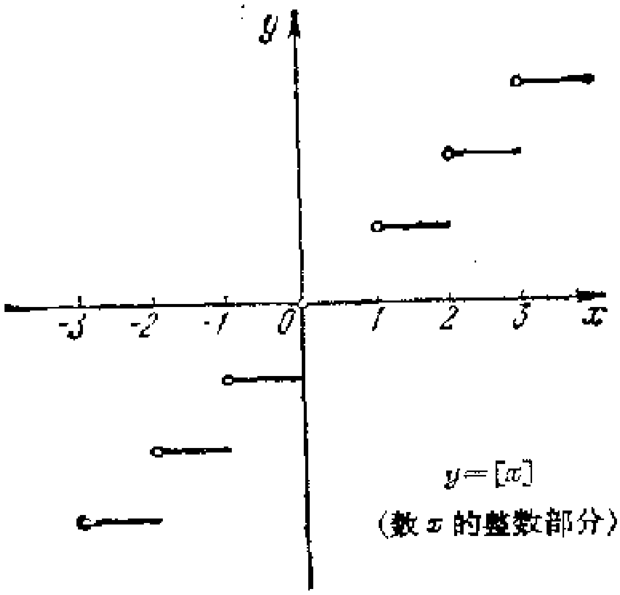


图9 d

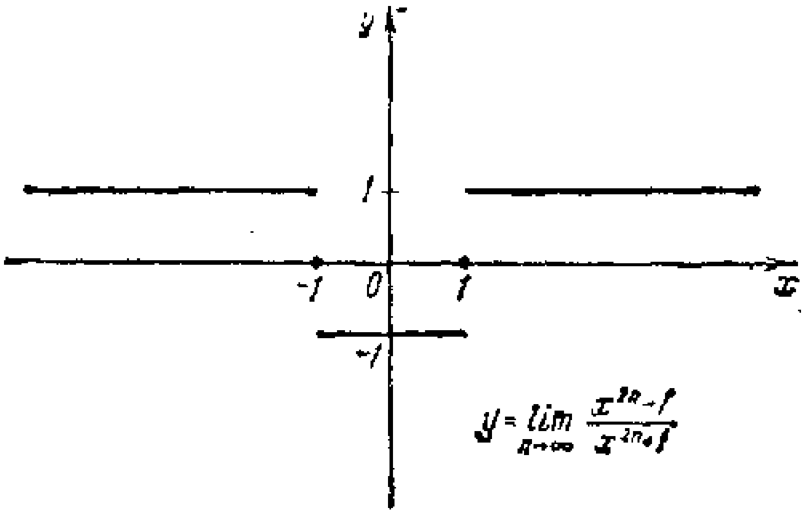


图9 e

$\frac{\sin x}{x}$, $\frac{1-\cos x}{x^2}$, $\frac{x^3-1}{x-1}$, $\frac{\operatorname{tg} x}{x}$ 在它们没有定义(就是分母变为零)的点上应该等于怎样的数，才使它们在这些点上连续，而对于函数 $\operatorname{tg} x$, $\frac{1}{x-1}$, $\frac{x-2}{x^2-4}$ 能不能规定出这样的数?

数学里的间断函数反映出在自然界中所碰到的许多飞跃式的过程．例如，在冲击的作用下，物体的速度就在数量上有飞跃的改

变．许多质的转变都伴有飞跃．在§2中我们曾引述了函数$Q=f(t)$的例子，这函数所表达的是一定数量的水(或冰)中所含的热量跟温度的相倚性．在临近冰的溶解温度的时候热量$Q=f(t)$随着t的变化而发生飞跃的改变．

具有个别间断点的函数，跟连续函数一样，是在分析中很频繁地会碰到的．至于更复杂的具有无穷多间断点的函数，我们可以引用所谓黎曼函数来作为一个例子．黎曼函数是这样一个函数：它在所有的无理点上等于零而在形如$x=\frac{p}{q}$(其中$\frac{p}{q}$是既约分数)的有理点上等于$\frac{p}{q}$．这函数在所有的有理点上间断而在所有的无理点上连续．把这函数稍微改动一下，就毫无困难地可以得到一个到处间断的函数的例子[1]．我们想顺便指出，甚至对于这样复杂的函数，现代的分析也发现了许多饶有趣味的规律性，在分析的独立分支之一——实变函数论——中就研究着这种规律性．苏联的数学家们，特别是莫斯科函数论学派，对于这一个在最近五十年来已有异常迅速发展的理论作出了巨大的贡献．

§5. 导　数

导数概念是分析的另一个主要概念．在历史上它是从下面两个问题的索解而产生的，就让我们先来考察一下这两个问题．

速度　在这一章的绪论中我们早已给自由落体的速度下了定义．那时我们利用取极限的步骤，从一小段路程上的平均速度求出了某一点上某一瞬时的速度．我们正可以根据同样的方法来定义任何非匀速运动中的瞬时速度．事实上，设函数

$$s=f(t) \tag{9}$$

所表达的是质点所经历的路程s跟时间t的相倚性．为了要得出在瞬时$t=t_0$的速度，让我们来考虑从t_0到$t_0+h(h\neq0)$的某一段时间．在这段时间内质点将经历路程

1) 只要让它在无理点上不等于零，比如说，等于1就够了．

$$\Delta s = f(t_0+h) - f(t_0).$$

在这段路程上的平均速度 v_{cp} 将随 h 而变：

$$v_{cp} = \frac{\Delta s}{h} = \frac{1}{h}\{f(t_0+h) - f(t_0)\},$$

若 h 愈小，则以 v_{cp} 表达瞬时 t_0 的真正速度就愈准确．由此可知，在瞬时 t_0 的真正速度等于路程增量与时间增量之比当后者始终不等于零而趋于零时的极限：

$$v = \lim_{h\to 0} \frac{f(t_0+h) - f(t_0)}{h}.$$

为了要算出在各种不同运动规律中的速度，我们还得学会怎样对各种不同的函数 $f(t)$ 去求出这个极限．

切线 另外一个问题，而且是几何上的问题——就是在任意平面曲线上求作切线的问题——也引导到完全相类似的极限的寻求．

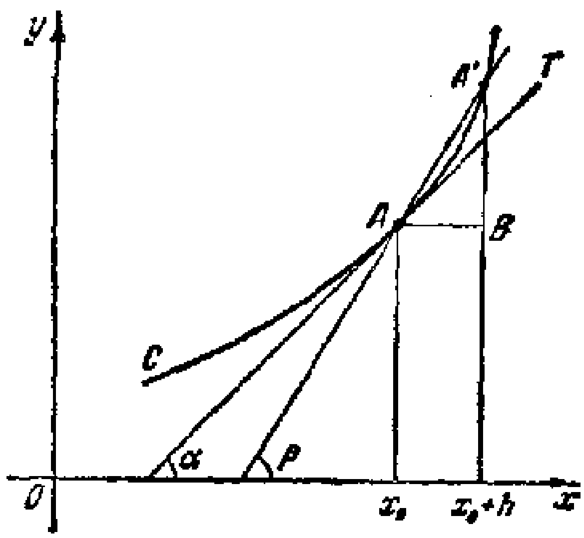


图 10

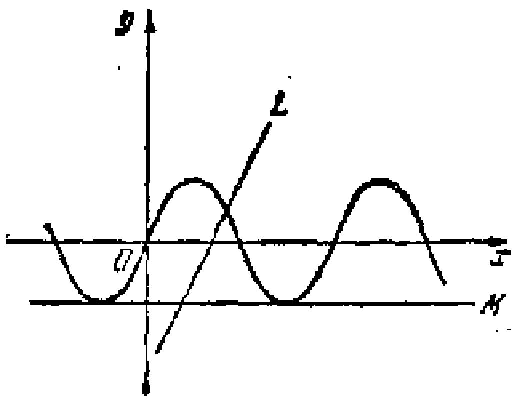


图 11

设曲线 C 是函数 $y=f(x)$ 的图形而 A 是 C 上具有横坐标 x_0 的一点(图 10)．怎样的直线称为曲线 C 在 A 点上的切线呢？在初等几何学里并未提出这个问题．对于在那里所研究的唯一曲线——圆线，我们把切线定义为跟圆周仅有一个公共点的直线．但是对于其他曲线，同样的定义显然将不符合"切线"的直觉概念了．例如，在图 11 中所绘的两条直线 L 和 M，前一条虽跟图中

的曲线(正弦曲线)只有一个公共点，显然不与曲线相切，而后一条跟曲线有许多公共点，却在每一个这样的点上与曲线相切.

为了要给切线下定义，让我们考虑曲线 C 上(图 10) 异于 A 而具有横坐标 x_0+h 的另外一点 A'. 作割线 AA' 并把它与 x 轴所成的夹角记作 β. 现在我们让点 A' 沿着曲线 C 趋近于 A. 若这时割线 AA' 趋于某一个极限位置，则占有这一极限位置的直线 T 就称为在 A 点上的切线. 显然，直线 T 与 x 轴所成的夹角 α 应该等于动角 β 的极限.

量 $\operatorname{tg}\beta$ 是易于由三角形 ABA'(图 10)来确定的：

$$\operatorname{tg}\beta=\frac{BA'}{AB}=\frac{f(x_0+h)-f(x_0)}{h}.$$

到了极限位置则应该有

$$\operatorname{tg}\alpha=\lim_{A'\to A}\operatorname{tg}\beta=\lim_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h},$$

这就是说，切线的倾斜角的正切等于函数 $f(x)$ 在 x_0 点上的增量与相应的自变量增量之比当后者始终不等于零而趋于零时的极限.

这里还有一个例子，也引导到类似的极限的寻求. 在导线中有一股强度变动不定的电流通过. 我们假设已经知道在时间 t 内流过导线某一固定横截面的电量由函数 $Q=f(t)$ 表出. 于是从 t_0 到 t_0+h 一段时间内将有等于 $f(t_0+h)-f(t_0)$ 的电量 ΔQ 流过这一横截面. 这时平均电流强度等于

$$I_{\text{cp}}=\frac{\Delta Q}{h}=\frac{f(t_0+h)-f(t_0)}{h}.$$

这一比式当 $h\to 0$ 时的极限就给出了在瞬时 t_0 的电流强度：

$$I=\lim_{h\to 0}\frac{f(t_0+h)-f(t_0)}{h}.$$

我们在上面所考虑的三个问题，虽然分属于人类知识的三个不同的领域：力学、几何学与电学，但是都引导到必须就某一函数来施行的同一种数学运算. 我们必须求出函数增量与相应的自变量增量 h 之比当 $h\to 0$ 时的极限. 我们尽可以随便再举出许多最不

相同的问题，如化学反应的速度问题，分布不均匀的质量的密度问题等等，而它们的求解都导致同样的运算．这种对函数所施行的运算，既然具有这样独特的作用，就得到了一个特别的名称——函数的微分运算．这种运算的结果则称为导数．

因此，函数 $y=f(x)$ 的导数，或者更准确些说，在给定一点 x 上的导数值是函数增量 $f(x+h)-f(x)$ 与自变量增量 h 之比当后者趋于零时的极限[1]．我们还往往使用这样的记号：$h=\Delta x$，$f(x+\Delta x)-f(x)=\Delta y$，于是导数的定义可简写为 $\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}$．

在 x 点上求得的导数值显然是随着 x 点而变的．所以函数 $y=f(x)$ 的导数又是 x 的某一个函数．导数通常是这样记写的：

$$f'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}.$$

我们还要指出其他通用的导数记号：

$$\frac{df(x)}{dx},\ \frac{dy}{dx},\ y'\ 或\ y'_x.$$

应当注意，记号 $\dfrac{dy}{dx}$ 虽然是当作导数的一个单独符号来念的，写出来却像分数的样子．在随后的几节中，这个"分数"的分子和分母对我们说来将获得独立的意义，而且它们的比式恰巧跟导数相等，于是显得这种写法是完全正确的了．

上面所探究的各个例子的结果现在可以重新叙述如下．

当动点所经历的路程 s 是时间的一个已知函数 $s=f(t)$ 时，动点的速度便等于这函数的导数：

$$v=s'=f'(t).$$

更简略地说，速度是路程对时间的导数．

在横坐标为 x 的点上作出曲线 $y=f(x)$ 的切线，其倾斜角的正切便等于函数 $f(x)$ 在这一点上的导数：

$$\operatorname{tg}\alpha=y'=f'(x).$$

1）当然这是就极限存在的场合而言．在相反的场合，则我们说在所考虑的 x 点上没有导数．

若 $Q=f(t)$ 所表达的是在时间 t 内流过导线横截面的电量，则在瞬时 t 的电流强度 I 便等于函数 $f(t)$ 的导数：

$$I=Q'=f'(t).$$

我们还要提出下面的说明．非匀速运动的瞬时速度是一个从实践中所产生的纯粹物理概念．人们是在对各种各样的具体运动进行了无数次的观察之后才达到这一概念的．物体在其路程的不同分段上所作的非匀速运动的研究，同时开始的各种不同的非匀速运动的比较，特别是物体碰撞现象的研究——这一切构成了非匀速运动的瞬时速度这一物理概念所由创造的实际资料．但是速度的确切定义一定必须本身包含有确定速度数量的方法．借助于导数概念这才可能做到．

在力学中，根据定义，若一个物体按照规律 $s=f(t)$ 运动，则其在瞬时 t 的速度数量就假定为等于函数 $f(t)$ 在 t 值上的导数．

这一节开头的推导，一方面表明了引用这种求导数的运算是有利的，而另一方面也给上述瞬时速度的定义找到了合理的根据．

因此，当我们刚提出在非匀速运动中寻求动点速度的问题时，实在说来，我们还只有关于速度数量的经验上的概念而并无确切的定义．由于相应的分析的结果，我们才达到了瞬时速度数量的确切定义．所得到的结果具有极重要的实际意义，因为在这个定义的基础上我们的关于速度的经验上概念更平添了把速度计算出来的可能性．

上面所说的话自然也适用于电流强度以及表示这种或那种(物理、化学等等)过程的速度的其他许多概念．

我们现在所指出的情况可以作为无数同类事实的范例：这就是实践引导到具有现实意义的一定概念(速度、功、密度、面积等等)，而数学使这概念获得清楚的定义，于是我们就有可能在我们所需要的计算中对所说的概念进行运算．

我们在这一章的开头早已指出了导数概念的发生首先是由于许多世纪以来竭力对曲线的切线及非匀速运动的速度两个问题进行探索的结果．类似的问题以及在后面将要讲到的面积计算问题

从古时候起就引起了数学家的注意．但直到十六世纪，这种问题的提法和解法还是带有非常特殊的性质．在这一方面所积累起来的丰富资料就在十七世纪——在牛顿和莱布尼茨的著作中——被系统化地整理而获得了理论上的完成．欧拉对于近世分析的奠基也作出了极巨大的贡献．

可是牛顿和莱布尼茨以及他们的同时代人很少在逻辑上给这一个伟大的数学发明确立根据；在他们的推论方法和所运用的概念中，可以找到许多在我们看来是模糊不清的东西；而且那时候的数学家们自己也意识到这一点，由他们在相互通信中对这些问题所进行的激烈争论可以作证．那时候（十七与十八世纪）的数学家们特别紧密地把自己的纯粹数学活动跟各种不同自然领域（物理学、力学、化学、技术）中的研究活动联系起来．数学问题的提出一般地是由于有实际的需要或愿望去考察这种或那种自然现象．在问题获得解决后，反正都得经受实际的考验，而正是这种情况有利地指引着数学的研究．

导数计算举例 我们既然把导数定义为极限

$$f'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h},$$

这就有可能对每一个具体的函数求出它的导数了．

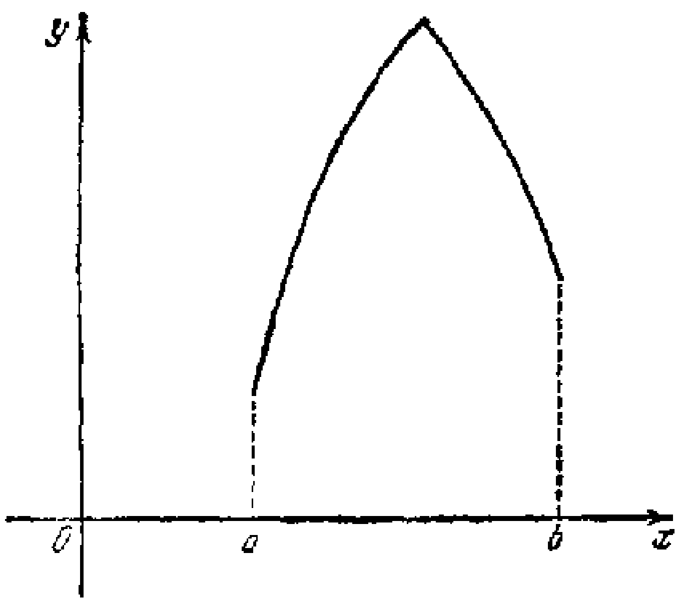


图 12

但是必须立刻说明，也可能有这样的场合，就是函数在这个或那个点上或甚至在许多点上根本没有导数，就是说，比式

$$\frac{f(x+h)-f(x)}{h}$$

当 $h\to 0$ 时干脆不趋于任何有限的极限．在函数的每一个间断点上显然就有这种场合，

因为这时，在比式

$$\frac{f(x+h)-f(x)}{h} \tag{10}$$

中，分子不趋于零而分母无限制地减小. 在函数连续的点上也可以没有导数. 函数图形上的折点(角点)就是一个简单的例子(图12). 试看图形中的曲线，在这样的点上没有确定的切线，相应于函数没有导数. 往往在这样的点上表达式(10)随着 h 自右或自左趋近于零而趋近于不同的值；但是若 h 任意地趋于零，则比式(10)根本没有极限. 导数不存在的更复杂一些的场合，可举这样一个函数作为例子：

$$y=\begin{cases} x\sin\dfrac{1}{x}, & \text{当 } x\neq 0, \\ 0, & \text{当 } x=0. \end{cases}$$

这函数的图形已在图13中绘出. 它在 $x=0$ 一点上没有导数，因为从图形中显然可见割线 OA，甚至当 A 从一方面趋近于零

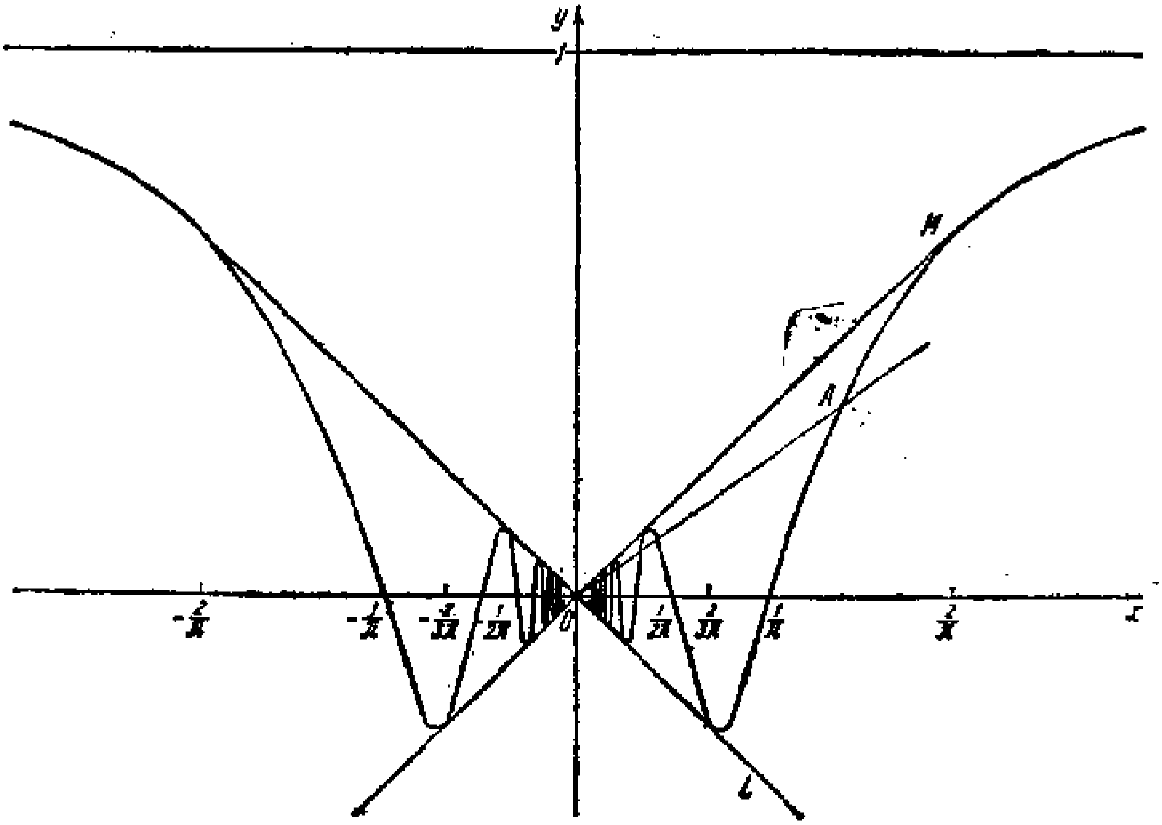


图 13

时，也不趋于确定的位置．这时割线 OA 始终在直线 OM 与直线 OL 之间摇摆着．在这一场合，比式(10)，甚至在 h 保持着正号或负号而趋于零时，也相应地没有极限．

最后，让我们指出，借助于公式我们还可以纯粹分析地给出在每一点上没有导数的连续函数．第一个提出这种函数的例子的是上一世纪的著名德国数学家魏尔斯特拉斯．

这样，把可微分[1]的函数归为一类，把一切连续函数也归为一类，则前一类大大地要比后一类来得狭隘．

现在让我们来实地计算一些最简单函数的导数．

1) $y=c$，式中 c 是常量．常量可以看作是函数的一种特殊场合，就是对于任何 x 它总等于同一个数．它的图形是一条平行于 x 轴而跟 x 轴保持着距离 c 的直线．这直线与 x 轴所成夹角为 $\alpha=0$，因此常量的导数显然恒等于零：$y'=(c)'=0$．由力学的观点看来，这等式表示不动点的速度等于零．

2) $y=x^2$．

$$\frac{f(x+h)-f(x)}{h}=\frac{(x+h)^2-x^2}{h}=2x+h.$$

当 $h\to 0$ 时[2]我们得到极限 $2x$，因而有

$$y'=(x^2)'=2x.$$

3) $y=x^n$（n 是正整数）．

$$\begin{aligned}\frac{f(x+h)-f(x)}{h}&=\frac{(x+h)^n-x^n}{h}\\&=\frac{1}{h}\left[x^n+nx^{n-1}h+\frac{n(n-1)}{2!}x^{n-2}h^2+\cdots+h^n-x^n\right]\\&=nx^{n-1}+\frac{n(n-1)}{2!}x^{n-2}h+\cdots+h^{n-1}.\end{aligned}$$

在最后一个等号右面的各项，从第二项起都在 $h\to 0$ 时趋于零，因此

$$y'=(x^n)'=nx^{n-1}.$$

1) 有导数存在称为可微分，简称可微．——译者注

2) 我们在这里处处认为 $h\neq 0$．

这公式，无论 n 是正的还是负的，是分数或甚至是无理数，也依旧正确，但是它的证明要用其他方法．这样，例如

$$(\sqrt{x})'=(x^{\frac{1}{2}})'=\frac{1}{2}x^{-\frac{1}{2}}=\frac{1}{2\sqrt{x}}\quad (x>0);$$

$$(\sqrt[3]{x})'=(x^{\frac{1}{3}})'=\frac{1}{3}x^{-\frac{2}{3}}=\frac{1}{3\sqrt[3]{x^2}}\quad (x\neq 0);$$

$$\left(\frac{1}{x}\right)'=(x^{-1})'=-1\cdot x^{-2}=-\frac{1}{x^2}\quad (x\neq 0);$$

$$(x^{\pi})'=\pi x^{\pi-1}\quad (x>0).$$

4) $y=\sin x$.

$$\frac{\sin(x+h)-\sin x}{h}=\frac{2\sin\frac{h}{2}\cos\left(x+\frac{h}{2}\right)}{h}=\frac{\sin\frac{h}{2}}{\frac{h}{2}}\cos\left(x+\frac{h}{2}\right).$$

但，我们在前面已经说明过，第一个分数在 $h\to 0$ 时趋于 1 而 $\cos\left(x+\frac{h}{2}\right)$ 显然趋于 $\cos x$．所以，正弦的导数等于余弦：

$$y'=(\sin x)'=\cos x.$$

我们想让读者用同样的推理自己来证明

$$(\cos x)'=-\sin x.$$

5) 在前面(第 103 页)我们已经指出下列极限的存在：

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n=e=2.71828\cdots,$$

前面也已经提到，在计算这极限时让 n 经历正整数值，这一情况并不起主要的作用，连续地趋向 ∞ 也可以．重要的是加在 1 后面的无穷小量 $\frac{1}{n}$，跟绝对值无限制增大的幂指数 n 是两个彼此互为倒数的量．

接受了这一个论断，我们就可以轻而易举地求到对数 $y=\log_a x$ 的导数．我们有

$$\frac{\log_a(x+h)-\log_a x}{h}=\frac{1}{h}\log_a\frac{x+h}{x}=\frac{1}{x}\log_a\left(1+\frac{h}{x}\right)^{\frac{x}{h}}.$$

由于对数的连续性，我们得以在极限中把符号 log 后面的量用它

的极限来替代，而这极限就等于 e:

$$\lim_{h\to 0}\left(1+\frac{h}{x}\right)^{\frac{x}{h}}=e.$$

(在这一场合，增大着的量 $\frac{x}{h}$ 就起着 $n\to\infty$ 的作用.）结果我们得到了对数的导数公式:

$$(\log_a x)'=\frac{1}{x}\log_a e.$$

这公式在取 e 这个数作为对数的底时变得特别简单. 以 e 作底的对数称为**自然对数**，并记作 $\ln x$. 于是我们可以写出

$$(\log_e x)'=\frac{1}{x},$$

或 $$(\ln x)'=\frac{1}{x}.$$

§6. 微分的法则

从上面所举的一些例子看来，似乎要计算一种新的函数的导数就需要创造一种新的方法. 但事实并不如此. 在不少程度上促成分析的发展的乃是这一情况：人们已经成功地创造出极简单而唯一的方法，可借以求出任何“初等”函数(就是那种可以由有限次基本代数运算、三角函数起乘幂及取对数来表达的函数)的导数. 作为这个方法的基础的就是所谓**微分的法则**. 这些法则是由一些化较复杂问题为较简单问题的定理所构成的.

我们将在这里表述微分的法则，而在推导时力求简短. 若读者想从这一章中得到的只是分析的一般概念，则尽可把这一节略去，但须记得确有方法去实地求出任何初等函数的导数. 在这种场合，读者对于将在后文举例中看到的一部分计算自然应当加以信赖.

和的导数 让我们假设 x 的函数 y 由下式给出:

$$y=\varphi(x)+\psi(x),$$

其中 $u=\varphi(x)$ 与 $v=\psi(x)$ 都是我们所已知的函数. 此外, 还假设我们能够求出函数 u 与 v 的导数. 那末, 怎样去求函数 y 的导数呢? 答案显得是简单的:

$$y'=(u+v)'=u'+v'. \tag{11}$$

事实上, 让我们给 x 以增量 Δx. 于是 u, v 与 y 也各得增量 Δu, Δv 与 Δy, 并有等式

$$\Delta y=\Delta u+\Delta v$$

成立. 据此[1],

$$\frac{\Delta y}{\Delta x}=\frac{\Delta u}{\Delta x}+\frac{\Delta v}{\Delta x};$$

取 $\Delta x\to 0$ 时的极限, 立即得公式(11), 只要函数 u 与 v 本身都确实有导数就是.

类似地可推得两函数之差的微分法则:

$$(u-v)'=u'-v'. \tag{12}$$

积的导数 乘积的微分法则稍为复杂一些. 两个可微函数之积也是可微函数, 其导数就等于第一个函数与第二个函数的导数之积, 再加上第二个函数与第一个函数的导数之积, 即

$$(uv)'=uv'+vu'. \tag{13}$$

事实上, 让我们给 x 以增量 Δx. 于是函数 u, v 及 $y=uv$ 分别得到增量 Δu, Δv, Δy, 而且这些增量满足关系式:

$$\Delta y=(u+\Delta u)(v+\Delta v)-uv=u\Delta v+v\Delta u+\Delta u\Delta v,$$

从而有

$$\frac{\Delta y}{\Delta x}=u\frac{\Delta v}{\Delta x}+v\frac{\Delta u}{\Delta x}+\Delta u\frac{\Delta v}{\Delta x}.$$

取 $\Delta x\to 0$ 时的极限, 等式右面的头两项就给出公式(13)的右面部分, 而第三项则告消失[2]. 所以从上面的等式取极限即得公式(13).

在特殊的场合, 若 $v=c=$ 常量, 则

---

1) 这里处处设 $\Delta x\neq 0$.

2) 最后一项在 $\Delta x\to 0$ 时趋于零, 因为 $\frac{\Delta v}{\Delta x}$ 趋于一个有限的数, 即一开始就假设为存在的导数 v', 而 $\Delta u\to 0$ (由于函数 u, 按照所设条件具有导数, 必然连续).

$$(cu)'=cu'+uc'=cu', \tag{14}$$

因为常量的导数等于零.

商的导数 设 $y=\frac{u}{v}$, 式中 u 与 v 在给定的 x 上都是可微函数, 而且就这些 x 而论 $v\neq 0$. 显然,

$$\Delta y=\frac{u+\Delta u}{v+\Delta v}-\frac{u}{v}=\frac{v\Delta u-u\Delta v}{(v+\Delta v)v},$$

从而有 $$\frac{\Delta y}{\Delta x}=\frac{v\frac{\Delta u}{\Delta x}-u\frac{\Delta v}{\Delta x}}{(v+\Delta v)v}\to\frac{vu'-uv'}{v^2}\quad(\Delta x\to 0).$$

在这里我们又用到了这一事实: 就可微函数 v 而论, 当 $\Delta x\to 0$ 时一定有 $\Delta v\to 0$. 因此,

$$\left(\frac{u}{v}\right)'=\frac{vu'-uv'}{v^2}. \tag{15}$$

下面是把所得到的法则应用起来的一些例子:

$$(2x^3-5)'=2(x^3)'-(5)'=2\cdot 3x^2-0=6x^2;$$

$$(x^2\sin x)'=x^2(\sin x)'+(x^2)'\sin x=x^2\cos x+2x\sin x;$$

$$(\operatorname{tg} x)'=\left(\frac{\sin x}{\cos x}\right)'=\frac{\cos x(\sin x)'-\sin x(\cos x)'}{\cos^2 x}$$

$$=\frac{\cos x\cdot\cos x-\sin x(-\sin x)}{\cos^2 x}$$

$$=\frac{1}{\cos^2 x}=\sec^2 x.$$

我们让读者自己来证明公式:

$$(\operatorname{ctg} x)'=-\operatorname{cosec}^2 x.$$

反函数的导数 让我们来考虑在区间 $[a, b]$ 上连续而且单调递增(或减)的函数 $y=f(x)$. 所谓单调递增(或减), 意思就是, 对应于区间 $[a, b]$ 上较大的 x 值的, 是较大(或较小)的 y 值(图 14).

设 $c=f(a)$ 而 $d=f(b)$. 在图 14 中显然可见, 对应于区间 $[c, d]$ (或 $[d, c]$) 上的每一个 y 值的, 是区间 $[a, b]$ 上满足 $y=f(x)$ 的某一个, 而且只有一个 x 值. 这样, 我们就在区间 $[c, d]$ (或 $[d, c]$) 上给出了一个完全确定的函数 $x=\varphi(y)$, 称为函数 $y=$

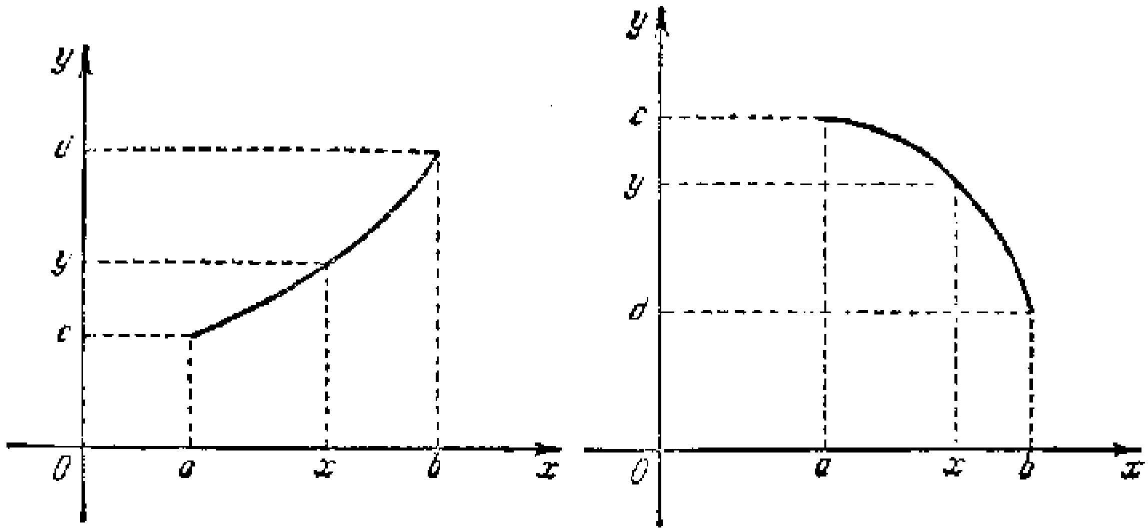


图 14

$f(x)$的反函数. 在图 14 中显然可见函数 $\varphi(y)$ 是连续的. 但这一事实在近世分析中是可以在分析的基础上严格地加以证明的. 现在设 Δx 与 Δy 是互相对应的 x 与 y 的增量. 显然若 $\Delta y \neq 0$, 我们就有

$$\frac{\Delta y}{\Delta x}=\frac{1}{\dfrac{\Delta x}{\Delta y}}.$$

取极限后, 这就给了我们正反两函数的导数之间的简单关系式:

$$y'_x=\frac{1}{x'_y}. \tag{16}$$

我们要利用这关系式来求函数 $y=a^x$ 的导数. 对于反函数 $x=\log_a y$ 我们已经知道怎样微分, 因此我们可以写出

$$(a^x)'_x=\frac{1}{(\log_a y)'_y}=\frac{1}{\dfrac{1}{y}\log_a e}=y\log_e a=a^x\ln a. \tag{17}$$

特别是

$$(e^x)'=e^x.$$

另外一个例子: $y=\arcsin x$. 反函数是 $x=\sin y$. 所以

$$(\arcsin x)'_x=\frac{1}{(\sin y)'_y}=\frac{1}{\cos y}=\frac{1}{\sqrt{1-\sin^2 y}}=\frac{1}{\sqrt{1-x^2}}.$$

导数表 让我们把一些最简单的初等函数的导数列表如下：

y	y'	y	y'	y	y'
c	0	$\ln x$	$\frac{1}{x}$	$\operatorname{tg} x$	$\sec^2 x$
x^a	ax^{a-1}	$\log_a x$	$\frac{1}{x}\log_a e$	$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$
e^x	e^x	$\sin x$	$\cos x$	$\arccos x$	$-\frac{1}{\sqrt{1-x^2}}$
a^x	$a^x \ln a$	$\cos x$	$-\sin x$	$\operatorname{arctg} x$	$\frac{1}{1+x^2}$

这些公式都是我们已经加以推导和阐释过的，只有最后两个是例外．这两个公式，读者如果愿意的话，可以自己用反函数的微分法则来加以推导．

复合函数的导数求法 我们还得看一看最后而且最难的一条微分法则．谁掌握了这条法则及导数表，我们就有充分的根据可以认为他能够对任何初等函数进行微分了．

为了要应用这一条我们即将在这里提出的法则，应当十分清楚地看到，需要加以微分的函数是怎样构成的，从自变量 x 出发必须经过哪些运算并且按照怎样的次序才终于得到函数 y．

例如，要计算函数

$$y=\sin x^2,$$

必须先把 x 平方，然后对所得到的量求正弦．这可以写成下面的样子：$y=\sin u$，其中 $u=x^2$．

反之，要计算函数

$$y=\sin^2 x,$$

必须先对 x 求正弦，然后把所得到的值平方．这可以写成：$y=u^2$，其中 $u=\sin x$．

这里还有一些例子：

1）$y=(3x+4)^3$；$y=u^3$，$u=3x+4$．

2）$y=\sqrt{1-x^2}$；$y=u^{\frac{1}{2}}$，$u=1-x^2$．

3）$y=e^{kx}$；$y=e^u$，$u=kx$．

在比较复杂的场合，可以得到一连串几个简单的相倚关系. 例如，

4) $y=\cos^3 x^2$, $y=u^3$, $u=\cos v$, $v=x^2$.

若 y 是变量 u 的函数：

$$y=f(u), \tag{18}$$

而 u 又是变量 x 的函数：

$$u=\varphi(x), \tag{19}$$

则 y，既是 u 的函数，也就是 x 的某一函数，记作

$$y=F(x)=f[\varphi(x)]. \tag{20}$$

把这一过程复杂化，还可以形成，比如说，相当于等式：

$$y=f(u),\ u=\varphi(v),\ v=\psi(x)$$

的函数
$$y=\Phi(x)=f\{\varphi[\psi(x)]\},$$
或者还可以得到更复杂的、归结为一连串同样等式的函数.

现在我们要说明怎样可把等式(20)所定义的函数 $F(x)$ 的导数计算出来，假设已经知道 $f(u)$ 对 u 的导数及 $\varphi(x)$ 对 x 的导数.

让我们给 x 以增量 Δx，于是由于等式(19)，u 将获得某一增量 Δu，而由于等式(18)，y 将获得增量 Δy. 我们可以写出

$$\frac{\Delta y}{\Delta x}=\frac{\Delta y}{\Delta u}\cdot\frac{\Delta u}{\Delta x}.$$

现在设 Δx 趋于零. 这时，$\frac{\Delta u}{\Delta x}\to u'_x$. 其次，由于 u 的连续性，增量 $\Delta u\to 0$，因而 $\frac{\Delta y}{\Delta u}\to y'_u$（导数 y'_u 与 u'_x 都假定是存在的）.

于是证明了关于复合函数求导数的重要公式[1]：

$$y'_x=y'_u u'_x. \tag{21}$$

让我们利用公式(21)及基本导数表（见第122页）来计算前面作为例子举出来的各函数的导数：

1) $y=(3x+4)^3=u^3$, $y'_x=(u^3)'_u(3x+4)'_x=3u^2\cdot 3$
$$=9(3x+4)^2.$$

1) 在推导这公式时，我们已经暗中假定，当 Δx 趋零时 Δu 始终不等于零. 事实上，即使在这假定不成立的场合，所说公式还是正确的.

2) $y=\sqrt{1-x^2}=u^{\frac{1}{2}}$, $y'_x=(u^{\frac{1}{2}})'_u(1-x^2)'_x$

$$=\frac{1}{2}u^{-\frac{1}{2}}(-2x)=-\frac{x}{\sqrt{1-x^2}}.$$

3) $y=e^{kx}=e^u$, $y'_x=(e^u)'_u\cdot u'_x=e^u\cdot k=ke^{kx}$.

若 $y=f(u)$, $u=\varphi(v)$, $v=\psi(x)$, 则

$$y'_x=y'_u\cdot u'_x=y'_u(u'_v\cdot v'_x)=y'_u\cdot u'_v\cdot v'_x.$$

怎样把这公式推广到有任意(有限)个接连成串的函数的场合，是显而易见的.

4) $y=\cos^3 x^2$, $y'_x=(u^3)'_u(\cos v)'_v(x^2)'_x=3u^2(-\sin v)2x$

$$=-6x\cos^2 x^2\sin x^2.$$

为了要说明怎样计算复合函数的导数，我们引用了中间变量 u, v, …，事实上，在经过一些训练后，就可以摆脱这种记号而只把它们记在心里.

初等函数 在结束这一节时，我们要指出，把前面导数表中所列出的那些函数作为基础，我们便可以给所谓初等函数下定义了．正是从这些最简单函数经过有限次的四则算术运算与函数复合运算而得到的一切函数称为**初等函数**.

例如，多项式 x^3-2x^2+3x-5 是初等函数，因为它是从形如 x^k 的一些函数借助于算术运算而得到的．函数 $\ln\sqrt{1-x^2}$ 也是初等的，因为它是从多项式 $u=1-x^2$ 借助于 $v=u^{\frac{1}{2}}$ 的运算及 $\ln v$ 的运算而得到的.

知道了最简单初等函数的导数而想寻求任何初等函数的导数，只要应用上述微分法则就足够了.

§7. 极大与极小．函数图形的研究

导数的最简单和重要的应用之一就是极大与极小理论．假设有函数 $y=f(x)$ 已在某一区间 $a\leqslant x\leqslant b$ 上给定．我们还假定这函

数不仅连续而且在所有的点上可微. 我们既然能够把导数算出, 也就有可能把函数图形的走势看出来. 在导数始终正值的分段上, 图形的切线是向上走的[1]. 在这样的分段上函数单调递增, 也就是说, 对应于较大的 x 值的是较大的 $f(x)$ 值. 反之, 在导数始终负值的分段上, 函数单调递减而图形向下走.

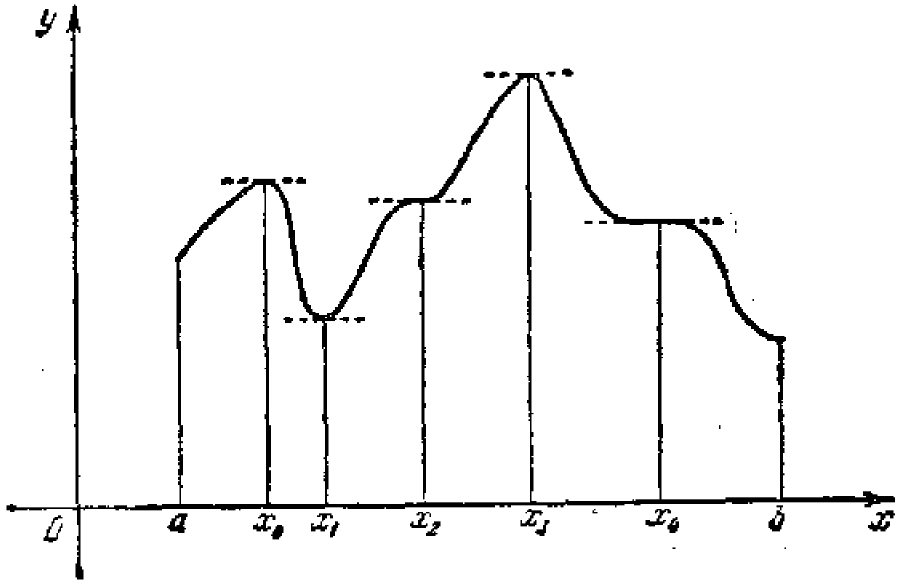


图 15

极大与极小 在图 15 中绘出了定义于区间 $[a, b]$ 上的函数 $y=f(x)$. 特别引人入胜的是图中具有横坐标 x_0, x_1, x_3 的点.

我们说函数 $f(x)$ 在 x_0 点上有一个局部的极大, 这样说就是要表示函数 $f(x)$ 在 x_0 点上的值大于邻点上的值, 或者更准确些说, 对于 x_0 点周围某一区间上的一切 x, $f(x_0)\geqslant f(x)$.

局部的极小是类似地来定义的.

就我们上面所绘的函数来说, 它在点 x_0 与 x_3 达到局部的极大, 在点 x_1 达到局部的极小.

在每一个极大或极小点上, 只要它是区间 $[a, b]$ 的内点, 就是说不跟端点 a 与 b 相重合的点, 导数必然等于零.

最后这个极重要的论断是从导数作为比式 $\frac{\Delta y}{\Delta x}$ 的极限这一定

1) 这是指在 x 增大的方向看切线地位的变动, 而切线的走势也就是图形的走势. ——译者注

义本身推出来的．事实上，从极大点稍稍移动地位，我们就有 $\Delta y \leqslant 0$，所以对于正的 Δx 比式 $\frac{\Delta y}{\Delta x}$ 是负的，而对于负的 Δx 比式 $\frac{\Delta y}{\Delta x}$ 是正的．这比式的极限，据假设是存在的，就不可能是正的或负的而只能是零．在直观上，这相应于下面的事实：在极大或极小点上（通常省去"局部"字样，但仍然包含有这样的意义）图形的切线是水平的．在图 15 中又可以看到，在点 x_2 与 x_4 上切线也是水平的，正像在点 x_0，x_1，x_3 上一样，可是在这两点上没有极大，也没有极小．使一个函数的导数等于零的那些点（逗留点）的数目，一般地说，可以多于极大与极小点的数目．

函数最大与最小值的求法　在非常不相同的各种技术问题中却都需要知道在怎样的 x 上这个或那个函数 $f(x)$ 达到了关于已给区间的最大或最小值．

在需要求出函数最大值的场合，问题就在于从区间 $[a, b]$ 上寻到一点 x_0，使得不等式 $f(x_0) \geqslant f(x)$ 对 $[a, b]$ 中的一切 x 都成立．

但是这里发生了一个原则性的问题：是不是一般地至少有一个这样的点存在呢？使用近世分析我们可以证明下面的存在定理：若函数 $f(x)$ 在有尽的闭区间上连续，则在这区间上至少存在着一点，使这函数达到关于区间 $[a, b]$ 的最大（或最小）值．

从上面所说的话我们可以推断：最大与最小点首先须在"逗留"点中去寻求．以此为基础就有了下面这个大家所知道的最大与最小值求法．

我们先求出 $f(x)$ 的导数，使其等于零，然后来求解所得到的方程：

$$f'(x)=0.$$

若 x_1，x_2，$\cdots$，x_n 是这方程的根，则可把 $f(x_1)$，$f(x_2)$，$\cdots$，$f(x_n)$ 各数互相比较一下．当然，我们还必须考虑到最大与最小值可能不在区间 $[a, b]$ 之内而在它的端点上（如图 15 中的最小值）或者在函数没有导数的点上（如图 12）出现．所以，在点 x_1，

$x_2, \cdots, x_n$ 之外，还要加上区间的端点 a 与 b 以及导数不存在的点，如果有的话．余下来的事就是把所有这些点上的函数值比较一下，在里边挑出最大的或最小的．

关于上述存在定理还得加上重要的补充．这就是，一般地说，当函数只是对于满足不等式 $a<x<b$ 的点集，即只是在开区间 (a, b) 内连续时，这定理就不再确实．我们让读者自己去考察函数 $\frac{1}{x}$ 在开区间 $(0, 1)$ 内既没有最大值，也没有最小值．

现在让我们来看一些例子．

从一块边长为 a 的正方形铁皮要做一个容积最大的无盖小方盒，若我们从铁皮的每一个角上割去边长为 x 的小方块（见 § 2，例 2），则得到盒的容积

$$V=x(a-2x)^2.$$

我们的问题归结于寻求那个在区间 $0\leqslant x\leqslant \frac{a}{2}$ 上使函数 $V(x)$ 达到最大值的 x．按照所说的法则，我们先求出导数并且让它等于零：

$$V'(x)=(a-2x)^2-4x(a-2x)=0.$$

解这个方程，我们得到两个根：

$$x_1=\frac{a}{2}, \quad x_2=\frac{a}{6}.$$

此外，还得加上函数 $V(x)$ 的定义区间的左端点（右端点跟 x_1 重合）．让我们比较在这些点上的函数值：

$$V(0)=0, \quad V\left(\frac{a}{6}\right)=\frac{2}{27}a^3, \quad V\left(\frac{a}{2}\right)=0.$$

于是，知道在高为 $x=\frac{x}{6}$ 时，盒将有最大容积，等于 $\frac{2}{27}a^3$．

作为第二个例子，让我们来看关于灯的问题（见 § 2，例 3）．要把溜冰道照得最亮，这个灯的高度 h 应该是多少？

由于 § 2 公式(3)，我们的问题归结于确定使 $T=\frac{A\sin\alpha}{h^2+r^2}$ 取得最大值的 h．但求 h 不如求角 α（见第 90 页图 3）来得方便．我

们有

$$h=r\operatorname{tg}\alpha,$$

因而
$$T=\frac{A}{r^2}\frac{\sin\alpha}{1+\operatorname{tg}^2\alpha}=\frac{A}{r^2}\sin\alpha\cos^2\alpha.$$

我们要在满足不等式 $0<\alpha<\frac{\pi}{2}$ 的 α 值中求出函数 $T(\alpha)$ 的最大值. 我们求出这函数的导数并且使其等于零:

$$T'(\alpha)=\frac{A}{r^2}(\cos^3\alpha-2\sin^2\alpha\cos\alpha)=0.$$

这方程可分解成为两个方程:

$$\cos\alpha=0,\ \cos^2\alpha-2\sin^2\alpha=0.$$

第一个方程有根 $\alpha=\frac{\pi}{2}$, 就是区间$\left(0,\ \frac{\pi}{2}\right)$的右端点, 第二个方程可化为

$$\operatorname{tg}^2\alpha=\frac{1}{2}.$$

但由于 $0<\alpha<\frac{\pi}{2}$, 得 $\alpha\approx35°15'$. 也正是这个 α 值使 $T(\alpha)$ 达到最大值(在区间的两端点上函数值较小, 因为在那里 $T=0$). 所求的高 h 将等于

$$h=r\operatorname{tg}\alpha=\frac{r}{\sqrt{2}}\approx0.7r.$$

所以要使溜冰道得到最好的照明, 灯应当安在近乎 $0.7r$ 的高度.

现在我们假定现有的设备还不能够把灯安装到比某一个 H 更大的高度. 这时角 α 的变化范围不可能是从 0 到 $\frac{\pi}{2}$, 而是比较狭隘的 $0<\alpha\leqslant\operatorname{arc\,tg}\frac{H}{r}$. 比如说, 设 $r=12$ 公尺, $H=9$ 公尺. 在这个场合, 事实上可以把灯装置在所需要的高度 $h=\frac{r}{\sqrt{2}}$, 约 8 公尺多一些. 但若 H 小于 8 公尺(比如说, 我们现有的装灯用的柱子只有 6 公尺长), 则函数 $T(\alpha)$ 在区间$\left[0,\ \operatorname{arc\,tg}\frac{H}{r}\right]$中没有等于零的导数. 在这个场合, 函数的最大值在区间的右端点上达到, 而灯就必须安装在现有的最大高度 $H=6$ 公尺.

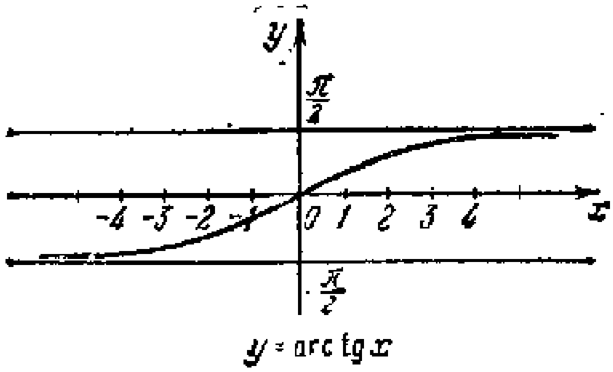


图 16 a

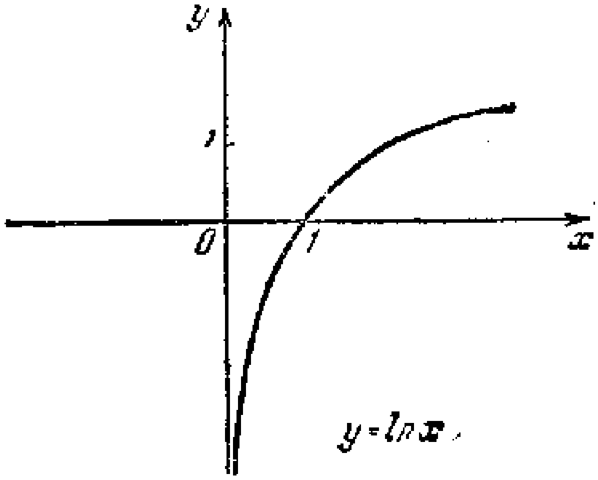


图 16 b

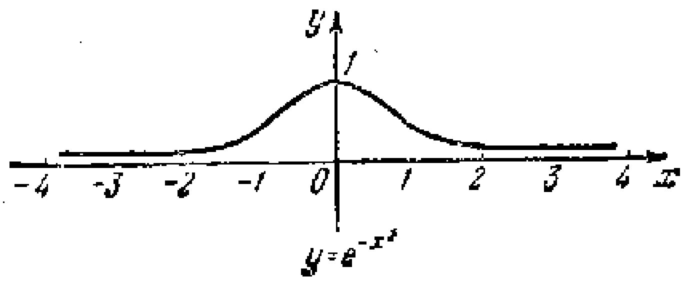


图 16 c

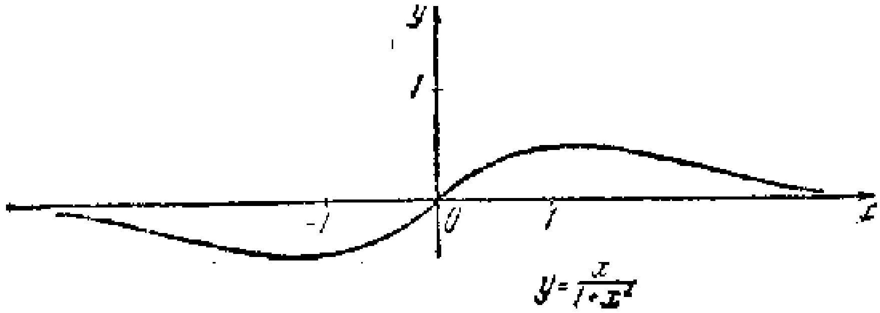


图 16 d

到现在为止，我们只考虑了有尽区间上的函数．若区间是无限的，则甚至于连续函数也可以在区间中达不到最大或最小值，而在 x 趋于无穷大时，比如说，始终增大或减小．

例如，函数 $y=kx+b$（见第93页图5），$y=\operatorname{arc\,tg} x$（图16 a），$y=\ln x$（图16 b）都达不到最大或最小值．函数 $y=e^{-x^2}$（图16 c）在点 $x=1$ 达到最大值，但没有最小值．至于函数 $y=\frac{x}{1+x^2}$（图16 d）它在点 $x=-1$ 达到最小值，在点 $x=1$ 达到最大值．

在区间无限的场合，最大与最小值的讨论仍可按照寻常的法则进行，只是我们必须考虑极限

$$A=\lim_{x\to-\infty} f(x),\ B=\lim_{x\to+\infty} f(x),$$

以替代 $f(a)$ 与 $f(b)$．

高阶导数 为了要比较周详地讨论函数 $f(x)$ 的图形，我们必需研究导数 $f'(x)$ 的变化进程．表达式 $f'(x)$ 本身又是 x 的某一个函数．对它我们也可以求导数．

导数的导数称为二阶导数，记作

$$[y']'=y'' \quad 或 \quad [f'(x)]'=f''(x).$$

类似地可以计算三阶导数：

$$[y'']'=y''' \quad 或 \quad [f''(x)]'=f'''(x)$$

等等，以及一般地，n 阶导数[1]：

$$y^{(n)}=f^{(n)}(x).$$

自然，我们必须注意到这个求导数的链锁可以在某一个 k 阶导数对于某些 x（或甚至对所有 x）中断；就是可以有导数 $f^{(k)}(x)$ 存在，而不再有导数 $f^{(k+1)}(x)$ 存在．任意阶导数我们将在下面 §9 考察泰勒公式时用到．现在我们所要讲的是二阶导数．

二阶导数的意义．曲线的凹凸 二阶导数有简单的力学意义．设 $s=f(t)$ 是点作直线运动的规律，于是 s' 就是速度，而 s'' 是“速度变化的速度”，或者更简单地说，点在瞬时 t 的加速度．例

---

1) 按照这个说法，最初的导数 $f'(x)$ 就称为一阶导数，而 $f(x)$ 也往往称为零阶导数．——译者注

如，在落体的场合，

$$s=\frac{gt^2}{2}+v_0t+s_0,$$

$$s'=gt+v_0,$$

$$s''=g,$$

即落体的加速度是常量.

二阶导数有简单的几何意义. 正像从一阶导数的正负号可以确定函数的增减，我们可以从二级导数的正负号判定函数图形中的曲线是向哪一面弯曲的.

若在某一分段上二阶导数始终是正的，则一阶导数单调递增，

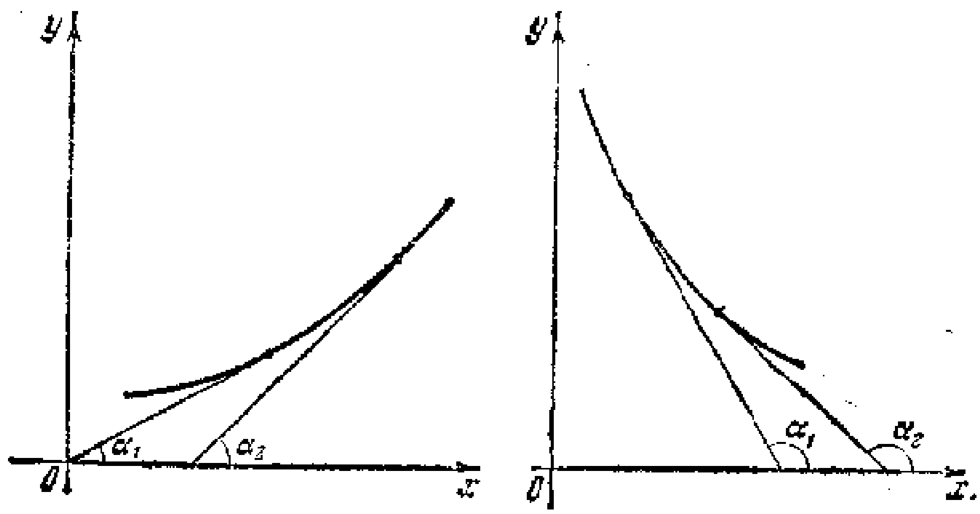


图 17

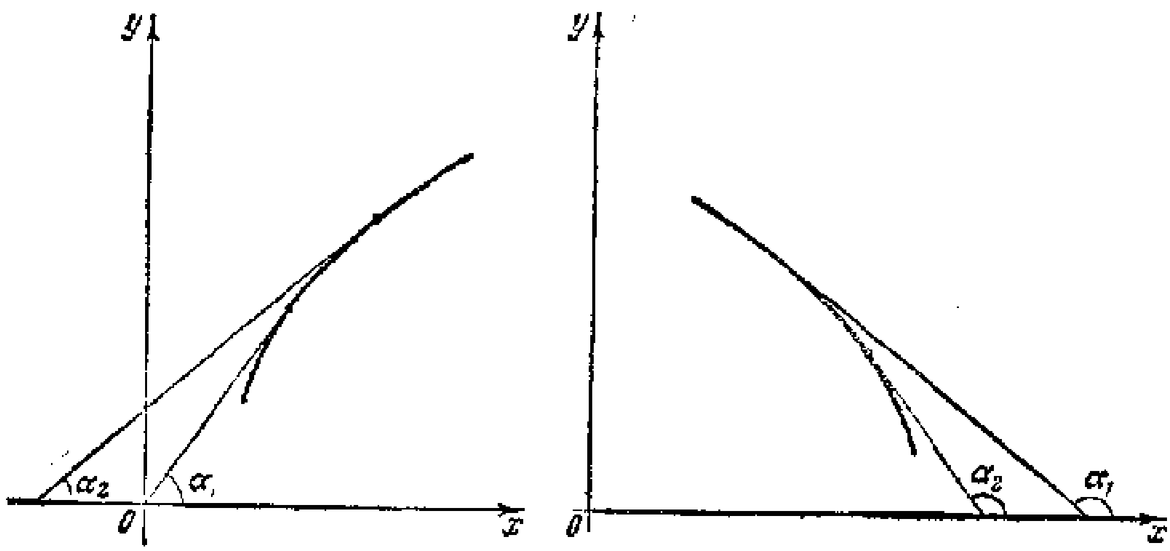


图 18

也就是 $f'(x)=\operatorname{tg}\alpha$ 单调递增，因此图形切线的倾斜角 α 本身也在增大(图 17)．这时沿着曲线移动直到下一分段的接界处为止，曲线是始终朝向同一面的，我们就说曲线是"向下凸"的．

反之，在二阶导数始终负值的分段上(图 18)函数图形的曲线是弯曲得"向上凸"的[1]．

极大与极小的准则．函数图形的研究 若曲线在整个所给定的 x 变化区间中是向上凸的，而在这区间的某一点 x_0 上有等于零的导数，则在这一点上曲线必然达到极大；在向下凸的场合则达到极小．由于这一个简单的推论，往往在找到了导数等于零的点后就可以弄清楚在这一点上是不是有局部的极大或极小[2]．

例 1 让我们来讨论函数

$$f(x)=\frac{x^3}{3}-\frac{5x^2}{2}+6x-2$$

的图形是怎么样的．

求出它的一阶导数并且让它等于零：

$$f'(x)=x^2-5x+6=0.$$

所得到的方程的根是 $x_1=2,\ x_2=3$．相应的函数值是

$$f(2)=2\frac{2}{3},\ f(3)=2\frac{1}{2}.$$

把所得到的两点绘在图上．还可以加上具有坐标 $x=0$ 与 $y=f(0)=-2$ 的点，这是图形与 y 轴的交点．二阶导数是 $f''(x)=2x-5$．它在 $x=\frac{5}{2}$ 时变为零，而且

$$\text{当}\quad x>\frac{5}{2}\ \text{时},\quad f''(x)>0,$$

$$\text{当}\quad x<\frac{5}{2}\ \text{时},\quad f''(x)<0.$$

---

1) 作为"向上凸"的严格定义的，是曲线的这一性质：连接曲线的任何两点作弦，曲线总在弦的上面(更准确些说：不在弦的下面)；类似地，在"向下凸"的时候，或者简短地说，在"凹"的时候，曲线就不在自己的弦的上面经过．

2) 在二阶导数本身也变号的较复杂场合，明确逗留点性质的问题是借助于泰勒公式(§9)来解决的．

点 $x=\frac{5}{2}$, $y=f\left(\frac{5}{2}\right)=2\frac{7}{12}$ 是图形的拐点. 在这一点的左面曲线向上凸, 右面向下凸.

现在显然可知, 点 $x=2$ 是函数的极大点而点 $x=3$ 是极小点.

根据所得到的数据, 我们就断定函数 $y=f(x)$ 的图形有着图 19 中所绘的那种样子. 曲线从点 $(0, -2)$ 起随着 x 的增大而增长, 并且是向上凸的, 在点 $\left(2, 2\frac{2}{3}\right)$ 达到它的极大, 然后转而向下. 在 $f''(x)=0$ 的点 $\left(2\frac{1}{2}, 2\frac{7}{12}\right)$, 凸一变而为凹. 其次, 曲线在点 $\left(3, 2\frac{1}{2}\right)$ 达到它的极小, 然后随着 x 的继续增大而向无穷大增长. 最后这个论断是由下面的事实推出来的: 函数中的第一项, 含有 x 的最高(三)次幂, 要比第二和第三项更迅速地趋于无穷大. 根据同样的理由, 当 x 取负值而绝对值增大时, 函数趋向 $-\infty$.

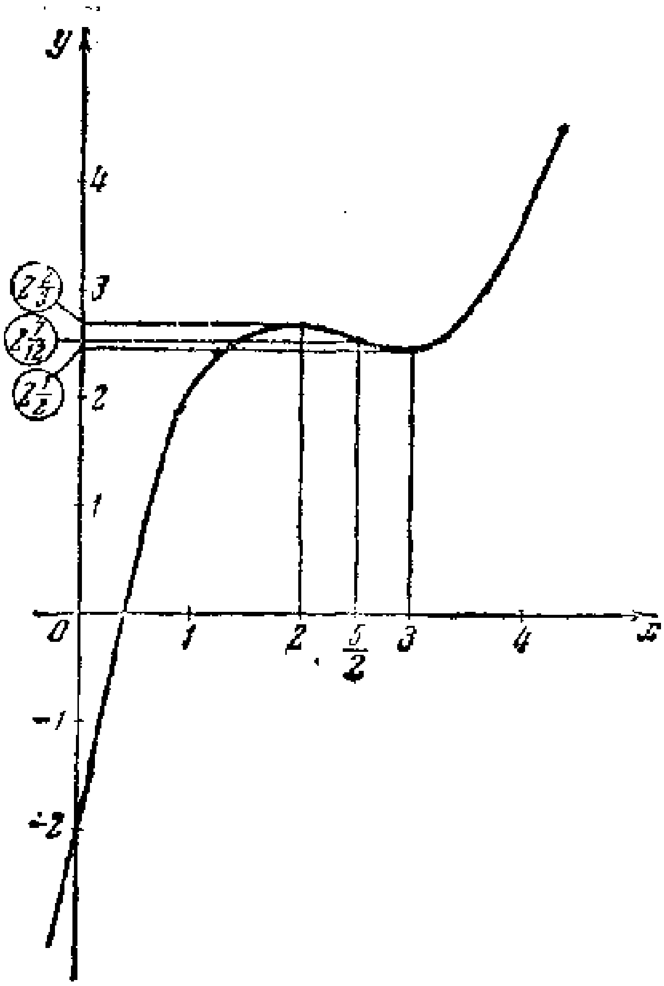


图 19

例 2 让我们证明对于任何 x 有不等式 $e^x \geqslant 1+x$ 成立. 为此, 我们要考察函数 $f(x)=e^x-x-1$. 它的一阶导数等于 $f'(x)=e^x-1$, 并且只在 $x=0$ 时变为零. 对于一切 x, 二阶导数 $f''(x)=e^x>0$. 所以函数的图形是向下凸的. 数 $f(0)=0$ 是函数的极小值, 因而对于一切 x, $e^x-x-1\geqslant 0$.

函数的讨论可以追求最不相同的各种目的. 例如, 借助于这

种讨论往往就可以查明这个或那个方程的实根个数．这样，为了要证明方程

$$xe^x=2$$

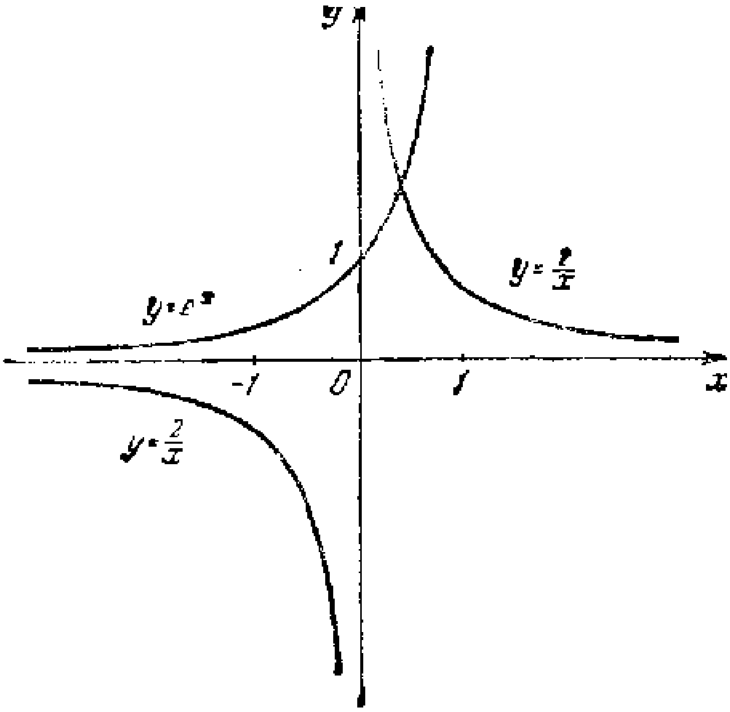


图 20

有唯一的实根，可以把函数 $y=e^x$ 及 $y=\frac{2}{x}$ 的图形(已在图 20 中绘出)讨论一下。容易看出，这两个函数的图形只相交于一点，因此方程 $e^x=\frac{2}{x}$ 只有一个根．

分析的方法被广泛地应用于方程的根的近似计算问题．关于这一问题见第四章 §5．

§8. 函数的增量与微分

函数的微分 让我们考虑可微函数 $y=f(x)$．相应于自变量增量 Δx 的函数增量

$$\Delta y=f(x+\Delta x)-f(x)$$

具有这样的性质：比式 $\frac{\Delta y}{\Delta x}$ 在 $\Delta x\to 0$ 时趋于有尽的极限，这极限就等于函数 $f(x)$ 的导数，即

$$\frac{\Delta y}{\Delta x}\to f'(x).$$

这个关系式可以写成等式的样子：

$$\frac{\Delta y}{\Delta x}=f'(x)+\alpha,$$

其中 α 是一个随 Δx 而定的量，而且在 $\Delta x\to 0$ 时它也趋于零．据此，函数增量有如下形式：

$$\Delta y=f'(x)\Delta x+\alpha\Delta x,$$

其中 $\alpha\to 0$，如果 $\Delta x\to 0$ 的话．

这等式的右面第一项是极简单地随着 Δx 而定的，即它是跟 Δx 成比例的．它就称为函数在已给定的 x 值上与自变量的已知增量 Δx 相对应的**微分**，记作

$$dy=f'(x)\Delta x.$$

前面那个等式的第二项具有这样的特征：由于里边有着因子 α，它在 $\Delta x\to 0$ 时比 Δx 更迅速地趋于零．我们就说，第二项是关于 Δx 的高阶无穷小，而在 $f'(x)\neq 0$ 的场合也是关于第一项的高阶无穷小．这样说，就是要表明在 Δx 充分小时第二项不仅本身很小，而且它跟 Δx 之比也变为任意地小．

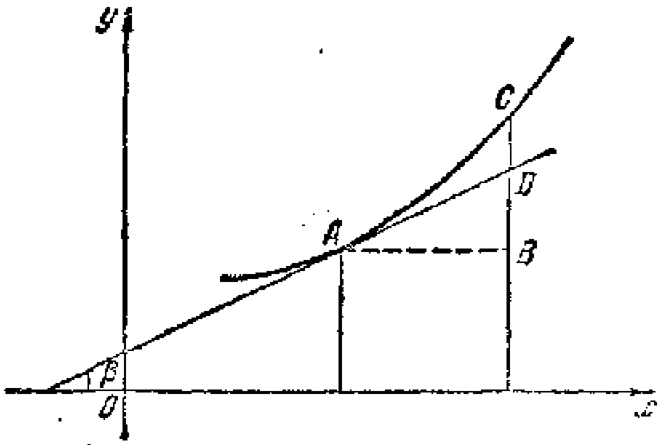


图 21

把 Δy 这样分解为两项，使其第一项（主要部分）线性地随 Δx 而变，其第二项在 Δx 微小时无足轻重地小，也可以在图 21 中看出来，线段 $BC=\Delta y$，同时 $BC=BD+DC$，其中 $BD=\operatorname{tg}\beta\cdot\Delta x=f'(x)\Delta x=dy$ 而 DC 是关于 Δx 的高阶无穷小．

在实践中往往利用微分作为函数增量的近似式．例如，设有一个有盖的立方小盒，盒内的长、宽、高各为 10 公分而盒壁厚 0.05 公分，欲求盒壁的体积．若我们在这里并不需要特别准确的结果，则可这样推论．所有盒壁的体积乃是函数 $y=x^3$ 在 $x=10$ 并相应于 $\Delta x=0.1$ 时的增量 Δy．近似地我们求得

$$\Delta y\approx dy=(x^3)'\Delta x=3x^2\Delta x=3\cdot 10^2\cdot 0.1=30\text{（立方公分）}.$$

为了记法的对称起见，自变量的增量 Δx 通常记作 dx，并且也

称为自变量的微分. 用了这样的记法，函数的微分将有如下的写法：

$$dy=f'(x)dx.$$

据此，导数就是函数微分与自变量微分之比：$f'(x)=\dfrac{dy}{dx}$.

函数的微分在历史上乃起源于“不可分量”概念. 这一个由现代观点看来远远不够清晰的概念在那时(在十七世纪)正是数学分析的基础. 关于这一概念的说法在几世纪中已经发生了实质上的变化. 不可分量，以及后来函数的微分，都曾经被当作真实的无穷小量——不知道怎样的好像是极微小的常量而同时并不是零. 在上面我们已经给出了在近世分析中所理解的微分的定义. 根据这个定义，对于自变量的每一个增量 Δx 说来，微分是有尽的量并且是跟 Δx 成比例的. 微分的另外一个主要性质——使它跟 Δy 区别开来的特征——只有在变化过程中才能够认识到，这就是：若我们考虑增量 Δx 趋于零(无穷小)，则 dy 与 Δy 之间的差将同时变为任意地小，即使是对 Δx 而言.

无穷小分析对于自然现象研究的大多数应用就建立在小增量可由微分来替代这一事实上. 这一点读者可从微分方程的实例看得特别清楚，本书的第二卷第五章和第六章就是专讲微分方程的.

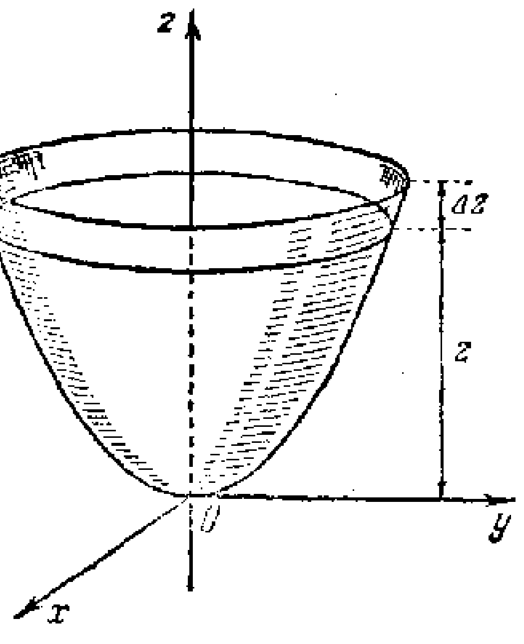


图 22

为了要知道表达某一过程的函数，我们先力图获得一个把这函数跟它的某阶导数确切地联系起来的方程. 把这种方程，就是所谓微分方程，建立起来的方法往往归结于把所求函数的增量由相应的微分来替代.

作为一个例子，让我们来

解下面的问题．在已给定了一个直角坐标系 $Oxyz$ 的空间，我们考虑一个曲面，由（在 Oyz 平面中）方程为 $z=y^2$ 的抛物线旋转而成．这曲面称为旋转抛物面（图 22）．设有一物体由这抛物面以及跟 Oxy 平面平行而且相距 z 的平面所围成，并以 v 记这物体的体积．显然，v 是 $z(z>0)$ 的一个函数．

为了要知道函数 v 等于什么，让我们来试求它的微分 dv．函数 v 在 z 点的增量 Δv 等于由抛物面及两个跟 Oxy 平面平行而且分别相距 z 与 $z+\Delta z$ 的平面所围成的体积．

容易看出，量 Δv 大于半径为 $\sqrt{z}$ 而高为 Δz 的圆柱体体积，但小于半径为 $\sqrt{z+\Delta z}$ 而高为 Δz 的圆柱体体积．这样，

$$\pi z\Delta z<\Delta v<\pi(z+\Delta z)\Delta z,$$

从而有

$$\Delta v=\pi(z+\theta\Delta z)\Delta z=\pi z\Delta z+\pi\theta(\Delta z)^2,$$

其中 θ 是随 Δz 而定并且满足不等式 $0<\theta<1$ 的某一个数．

这样，我们已成功地将增量 Δv 表达为和的形式，其第一项与 Δz 成比例而第二项为关于 Δz（当 $\Delta z\to 0$ 时）的高阶无穷小．由此可知第一项就是函数 v 的微分：

$$dv=\pi z\,\Delta z,$$

或

$$dv=\pi z\,dz,$$

因为对于自变量 z 有等式 $\Delta z=dz$ 成立．

所得到的等式把（变量 v 与 z 的）微分 dv 与 dz 互相联系起来，所以称为**微分方程**．

若我们注意到

$$\frac{dv}{dz}=v',$$

其中 v' 是 v 对变量 z 的导数，则我们的微分方程还可以写成下面的形式：

$$v'=\pi z.$$

要解出这个极简单的微分方程就是要寻出其导数是等于 πz 的 z 的函数．在后面 §10 与 §11 中将以普遍形式讲到同样的问题，但现在我们让读者自己去验证上述问题的解是函数

$v=\frac{\pi z^2}{2}+C$, 其中 C 可以取作任何常数[1]. 在这一场合，函数所表达的是我们所考虑的物体的体积，显然在 $z=0$ 时等于零(看图 22)，因此 $C=0$, 所以我们的函数就由等式 $v=\frac{\pi z^2}{2}$ 确定.

中值定理及其应用的举例 微分是通过自变量的增量与起点上的导数来表达函数增量的近似值的. 若问题在于从 $x=a$ 到 $x=b$ 一段上的增量, 则有

$$f(b)-f(a)\approx f'(a)(b-a).$$

我们也可以得到这种样子的准确的等式，只要把右面起点上的导数 $f'(a)$ 用区间 (a, b) 中适当地选定的某一点上的导数来替代就是. 更准确些说：若 $y=f(x)$ 是区间 $a\leqslant x\leqslant b$ 上的可微函数，则在区间的内部必有一点 ξ, 使准确的等式

$$f(b)-f(a)=f'(\xi)(b-a) \tag{22}$$

成立.

这条“中值定理”，又称为拉格朗日公式或有限增量公式，具有极简单的几何意义. 设在函数 $f(x)$ 的图形中把对应于 $x=a$ 与 $x=b$ 的两点 A 与 B 用弦 AB 连接起来(图 23). 让我们把直线 AB 跟本身平行地向上或向下移动. 于是当我们的直线跟图形最

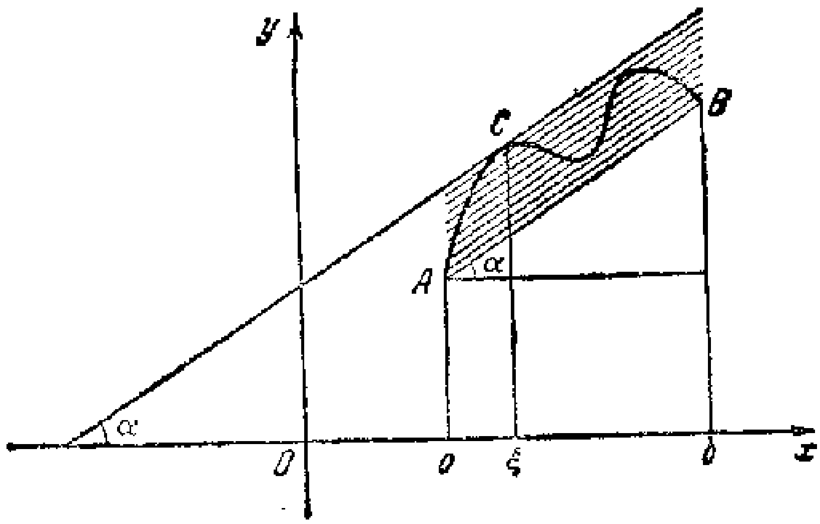


图 23

1) 这就给出了一切的解(见第 149 页注解).

后一次相交的时候，这直线将跟图形相切于某一点 C. 在这一点(设其横坐标 $x=\xi$)上，切线跟弦 AB 有同样的倾斜角 α. 但对弦说来，

$$\operatorname{tg}\alpha=\frac{f(b)-f(a)}{b-a}.$$

另一方面，在 C 点上

$$\operatorname{tg}\alpha=f'(\xi).$$

等式
$$\frac{f(b)-f(a)}{b-a}=f'(\xi)$$
正好表达了中值定理[1].

公式(22)有一个独特的地方，就是在它里边出现了一个对我们说来是未知的点 ξ，我们所知道的每次只是它在"区间 (a, b) 的某处". 尽管具有这样的不确定性，这公式还是有重大的理论意义，是分析中许多定理赖以证明的工具. 它的直接的实践意义也是巨大的，因为它使得有可能在导数的变动范围为已知时把函数的增量估计出来. 例如，

$$|\sin b-\sin a|=|\cos\xi|(b-a)\leqslant b-a.$$

在这里 a, b 及 ξ 都是以弧度表出的角；ξ 是 a 与 b 之间的某一个值；它是未知的，但我们知道 $|\cos\xi|\leqslant 1$.

由公式(22)显然可知，导数始终等于零的函数必为常量，它在任何分段上不可能得到异于零的增量. 通过类似的途径读者现在容易证明，导数始终为正的函数必然单调增，而在导数为负时单调减. 让我们不加证明提出中值定理的一种推广：

对于任何在 $[a, b]$ 中可微的函数 $\varphi(x)$ 与 $\psi(x)$[只须在 (a, b) 内 $\psi'(x)\neq 0$]，下列等式是确实的[2]：

$$\frac{\varphi(b)-\varphi(a)}{\psi(b)-\psi(a)}=\frac{\varphi'(\xi)}{\psi'(\xi)},\tag{23}$$

1) 当然，上述推论只是说明了定理的几何意义，但并不是定理的严格证明.

2) 直接地应用中值定理于函数

$$f(x)=\varphi(x)-\frac{\varphi(b)-\varphi(a)}{\psi(b)-\psi(a)}\psi(x),$$

即可得到公式(23).

其中 ξ 是区间 (a, b) 的某一点[1].

由这一论断可以得到一种普遍的方法来计算如下形式的极限:

$$\lim_{x\to 0}\frac{\varphi(x)}{\psi(x)}, \tag{24}$$

其中 $\varphi(0)=\psi(0)=0$. 利用公式(23), 让我们指出

$$\frac{\varphi(x)}{\psi(x)}=\frac{\varphi(x)-\varphi(0)}{\psi(x)-\psi(0)}=\frac{\varphi'(\xi)}{\psi'(\xi)},$$

其中 ξ 位于 0 与 x 之间, 因此跟 x 一起趋于 0. 这使我们可以用 $\lim\limits_{x\to 0}\dfrac{\varphi'(x)}{\psi'(x)}$ 来替代极限(24), 在许多场合就大大地便利了极限的寻求[2].

例 试求 $\lim\limits_{x\to 0}\dfrac{x-\sin x}{x^3}$. 三次应用所说的法则, 我们依次得到

$$\lim_{x\to 0}\frac{x-\sin x}{x^3}=\lim_{x\to 0}\frac{1-\cos x}{3x^2}=\lim_{x\to 0}\frac{\sin x}{6x}=\lim_{x\to 0}\frac{\cos x}{6}=\frac{1}{6}.$$

§9. 泰勒公式

函数

$$p(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$$

(其中系数 a_k 都是常数)称为 n 次多项式. 特别是函数 $y=ax+b$ 与 $y=ax^2+bx+c$ 分别为一次与二次多项式. 多项式可以看作是最简单的函数. 按已给定的 x 值计算多项式, 用加、减、乘三种运算就已足够, 连除法也是不需要的. 多项式对于任何 x 是连续的并有任意阶的导数. 顺便指出, 多项式的导数还是多项式, 但次数减少一, 而 n 次多项式的 $n+1$ 阶及更高阶的导数就都等于零了.

---

1) 关于$[a, b]$与(a, b)这两个记号参阅第 91 页注解. ——译者注

2) 在寻求分子与分母都趋于无穷大的比式的极限时, 这一法则还是正确的. 这方法对于寻求这样的极限(又称为解不定型)极有效用, 以后, 比如说, 在第二卷第十二章 §3 中也将加以应用.

若在多项式之外再加上如下形式的函数：

$$y=\frac{a_0+a_1x+\cdots+a_nx^n}{b_0+b_1x+\cdots+b_mx^m},$$

其计算已经需要除法，还加上函数 $\sqrt{x}$ 与 $\sqrt[3]{x}$，以及最后加上这种种函数的算术组合——这些就是我们借助于中学教程里所学到的方法就能够计算的一切函数.

我们在中学时也已经得到了关于其他许多函数的概念，如

$$\sqrt[5]{x},\ \lg x,\ \sin x,\ \operatorname{arc\,tg} x,\ \cdots$$

我们知道了这些函数的最重要性质，但是初等数学不曾回答这一个问题：怎样来计算它们？例如，要得到 $\lg x$ 或 $\sin x$，必须对 x 作怎样的运算？对于这种问题，分析中所研究出来的方法才提供了答案．我们在这里要比较详细地说明其中的一种方法.

泰勒公式 设在某一个里边包含有 a 点的区间上给定了一个任意阶可微的函数 $f(x)$．一次多项式

$$p_1(x)=f(a)+f'(a)(x-a)$$

就在 $x=a$ 点上跟 $f(x)$ 相合，而且不难验证在这一点上还跟 $f(x)$ 有同样的导数．这多项式的图形是一条跟 $f(x)$ 的图形在 a 点相切的直线．我们又可以挑出这样一个二次多项式，就是

$$p_2(x)=f(a)+f'(a)(x-a)+\frac{f''(a)}{2}(x-a)^2,$$

使在 $x=a$ 点上跟 $f(x)$ 有共同的值以及同样的一阶与二阶导数．这二次多项式的图形将在 a 点的附近更密切地紧贴着函数 $f(x)$ 的图形．自然我们希望，假如我们造出一个多项式，在 $x=a$ 上跟 $f(x)$ 有同样的一阶至 n 阶导数，那末这多项式在与 a 邻近的 x 值上将更好地接近于 $f(x)$．这样就得到了表达泰勒公式的近似等式：

$$f(x)\approx f(a)+f'(a)(x-a)+\frac{f''(a)}{2!}(x-a)^2+\cdots+\frac{f^{(n)}(a)}{n!}(x-a)^n. \tag{25}$$

这公式的右面是 $x-a$ 的 n 次多项式．若已知 $f(a), f'(a), \cdots,$

$f^{(n)}(a)$，我们就可以对每一个 x 实地来计算这个多项式．

对于 $n+1$ 阶可微的函数，容易证明，这公式的右面跟左面只相差一个微小的比 $(x-a)^n$ 更迅速地趋于零的量．此外，这是唯一可能的 n 次多项式，它在与 a 邻近的 x 点上跟 $f(x)$ 相差一个当 $x\to a$ 时比 $(x-a)^n$ 更迅速地趋于零的量．在 $f(x)$ 是 n 次代数多项式的场合，近似等式(25)就变为准确的等式．

最后，而且异常重要的，我们还能够简单地表出究竟公式(25)的右面跟 $f(x)$ 相差多少．这就是说，为了要使等式(25)成为准确的等式，还必须在右面加上这公式的所谓“余项”：

$$f(x)=f(a)+f'(a)(x-a)+\cdots+\frac{f^{(n)}(a)}{n!}(x-a)^n$$
$$+\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}. \tag{26}$$

添上去的最后一项[1)]

$$R_{n+1}(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$$

具有这样的特性，就是这里边的导数应当每次不在 a 点本身上，而在特别挑定的、我们并不预先知道的、但总在 a 与 x 所成区间之内的一点 ξ 上来计算．

等式(26)的证明虽颇繁重，但在实质上并不太复杂．让我们来引述一种稍微有点矫揉造作的而却简短的证明．

为了要确定在近似等式(25)中左面跟右面究竟相差多少，让我们来考虑等式(25)左右两面的差跟量 $-(x-a)^{n+1}$ 之比：

$$\frac{f(x)-\left[f(a)+f'(a)(x-a)+\cdots+\frac{f^{(n)}(a)}{n!}(x-a)^n\right]}{-(x-a)^{n+1}}. \tag{27}$$

我们还要引用变量 u 的函数

$$\varphi(u)=f(u)+f'(u)(x-u)+\cdots+\frac{f^{(n)}(u)}{n!}(x-u)^n,$$

其中 x 当作是固定的量（即常量）．于是(27)式的分子不是别的，

1) 这只是余项 $R_{n+1}(x)$ 的各种可能表达式中的一种．

而是函数 $\varphi(u)$ 从 $u=a$ 转变到 $u=x$ 时的增量，而分母是函数

$$\psi(u)=(x-u)^{n+1}$$

在同一分段上的增量．余下来的事就是利用在上一节中所说的中值定理的推广：

$$\frac{\varphi(x)-\varphi(a)}{\psi(x)-\psi(a)}=\frac{\varphi'(\xi)}{\psi'(\xi)}.$$

把函数 $\varphi(u)$ 与 $\psi(u)$ 对 u 求导数（这时必须记得 x 是常量；我们已经把它固定起来），就可知

$$\frac{\varphi'(\xi)}{\psi'(\xi)}=-\frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

令最后的表达式跟起初的量(27)相等，恰巧得到(26)那种样子的泰勒公式．

从(26)那种样子的泰勒公式不仅仅是近似地计算 $f(x)$ 的工具，而且还使我们得以估计出这里所容许的误差．

现在来看一个简单的例子：

$$y=\sin x.$$

当 $x=0$ 时函数 $\sin x$ 及其任意阶导数的值都是我们知道的．让我们利用这一情况，把 $\sin x$ 的泰勒公式写出，假设 $a=0$ 并以 $n=4$ 的场合为限．我们逐一求得

$$\begin{aligned}
&f(x)=\sin x, && f'(x)=\cos x, && f''(x)=-\sin x,\\
&f'''(x)=-\cos x, && f^{\mathrm{IV}}(x)=\sin x, && f^{\mathrm{V}}(x)=\cos x,\\
&f(0)=0, && f'(0)=1, && f''(0)=0,\\
&f'''(0)=-1, && f^{\mathrm{IV}}(0)=0, && f^{\mathrm{V}}(\xi)=\cos\xi.
\end{aligned}$$

所以 $$\sin x=x-\frac{x^3}{6}+R_5,\quad \text{式中}\ R_5=\frac{x^5}{120}\cos\xi.$$

虽然准确的 R_5 值是我们所不知道的，但是由于 $|\cos\xi|\leqslant 1$，我们不难加以估计．若把 x 的值限于从 0 到 $\frac{\pi}{4}$，则对于这样的 x 有

$$|R_5|=\left|\frac{x^5}{120}\cos\xi\right|\leqslant\frac{1}{120}\left(\frac{\pi}{4}\right)^5<\frac{1}{400}.$$

因此，在区间 $\left[0,\ \frac{\pi}{4}\right]$ 上函数 $\sin x$ 可以在 $\frac{1}{400}$ 的准确度内当作是

等于三次多项式

$$\sin x = x - \frac{1}{6}x^3.$$

若我们在 $\sin x$ 按泰勒公式展开时取更多的项，则可以得到更高次的多项式，更准确地与 $\sin x$ 相近似。

三角及其他许多算表都是用同样的方法计算出来的。

自然界的规律一般地可以很近似地由任意阶可微的函数来表达，而这种函数又可近似地由多项式表出；多项式次数的选择决定于所需要的准确度。

泰勒级数 若在公式(25)中所取的项数愈来愈多，则由余项 $R_{n+1}(x)$ 所表达的(25)式右面跟 $f(x)$ 的误差可以趋近于零。当然，这决不是经常的情况：并非对每一个函数，对每一个 x 值都如此。但是有广阔的一类函数(称为分析函数)，其余项至少对于包围着点 a 的某一区间的 x 值确实在 $n\to\infty$ 时趋于零。正是对这样的函数说来，泰勒公式使我们可以把 $f(x)$ 计算到任何准确度。让我们更仔细地来看一看这种函数。

若在 $n\to\infty$ 时 $R_{n+1}(x)\to 0$，则由(26)得

$$f(x) = \lim_{n\to\infty}\left[f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n\right].$$

在这一场合，我们说，$f(x)$ 已展开为按 $x-a$ 的升幂排列的收敛无穷级数：

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots$$

称为泰勒级数，而且 $f(x)$ 就称为这级数的和。让我们举一些大家所熟知的函数被展开为泰勒级数的实例(在这些例子中 $a=0$)：

1) $(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots$

(对 $|x|<1$ 及任何实数 n 正确)。

2) $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$ (对一切 x 正确)。

3) $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$ (对一切 x 正确)。

4) $e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots$　　(对一切 x 正确).

5) $\operatorname{arc\,tg} x=x-\frac{x^3}{3}+\frac{x^5}{5}-\cdots$　　(对 $|x|\leqslant 1$ 正确).

这些例子中的第一个就是著名的牛顿二项式，首先由牛顿求出对一切 n 有效，但在他那时候还完全只以整数的 n 作为根据．这一例子成了后来建立普遍泰勒公式的范例．最后两个例子则在 $x=1$ 时可用来计算数 e 和 π 到任何准确度．

泰勒公式给分析应用中的大多数计算问题开辟了道路，其实践的意义是非常巨大的．

被展开为泰勒级数的函数以很大的准确度表达了自然界中的许多规律性：物理与化学的过程，物体的运动等等．只要把它们考虑作复变量函数，它们的理论就获得充分的明晰性和完美性．这种函数的理论将在第二卷第九章中占有相应的地位．

函数由多项式近似地来表达这一概念本身以及用无穷多个较简单项之和的形式来表示函数的问题在分析中得到了远大的发展，成为数学的一个独立分支——函数逼近理论(见第二卷第十二章)．

§10. 积　分

从第一章以及本章的 §1 读者早已知道，积分的概念及一般地说积分学乃起源于需要求解某些具体的问题，其典型的例子是寻求曲线形面积的问题．这一节所要讲的就是这个问题．由这个问题我们也将知道微分学与积分学问题之间的联系究竟在哪里，这在十八世纪才完全弄清楚．

面积　设在 x 轴之上有一条曲线作为函数 $y=f(x)$ 的图形．让我们试求由曲线 $y=f(x)$，x 轴以及通过点 $x=a$ 与 $x=b$ 而跟 y 轴平行的两条直线所围成的一块面积 S．

要解决这个问题，让我们象下面这样子来做．把区间 $[a, b]$ 分成 n 个(不必等长的)小段，第一个小段的长度记作 Δx_1，第二个

$\Delta x_2, \cdots$，最后一个 Δx_n. 在每一个小段中各挑定一点 $\xi_1, \xi_2, \cdots, \xi_n$，

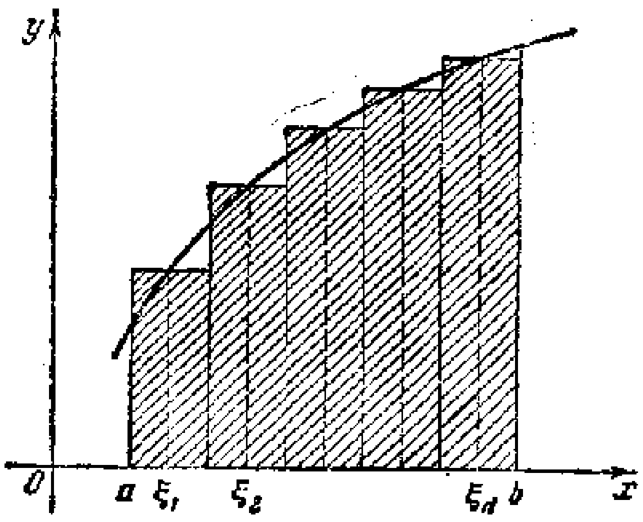


图 24

然后让我们组成下面的和：

$$S_n = f(\xi_1)\Delta x_1 + f(\xi_2)\Delta x_2 + \cdots + f(\xi_n)\Delta x_n. \quad (28)$$

量 S_n 显然等于图 24 中加有阴影的各矩形的面积之和.

把区间 $[a, b]$ 分得愈细，S_n 将跟面积 S 愈接近. 若我们把 $[a, b]$ 分成愈来愈细的小段而作出一连串这样的图，则和 S_n 将趋于 S.

因为 $[a, b]$ 是可以分作不相等的小段的，我们还得明确一下所谓分得"愈来愈细"应当怎样理解. 我们要假定，不仅 n 无限制增大，而且甚至连各小段长度 Δx_i 中的最大者也趋于零. 于是

$$S = \lim_{\max \Delta x_i \to 0} [f(\xi_1)\Delta x_1 + f(\xi_2)\Delta x_2 + \cdots + f(\xi_n)\Delta x_n]$$

$$= \lim_{\max \Delta x_i \to 0} \sum_{i=1}^{n} f(\xi_i)\Delta x_i. \quad (29)$$

面积的计算已归结为极限(29)的寻求.

我们要指出，当我们提出我们的问题时，我们还只有关于曲线形面积的经验上概念，而并没有确切的定义. 由于上面所作的推论的结果，我们才获得了面积概念的确切定义. 这就是极限(29). 现在我们不仅有面积的直觉概念，而且还拥有它的数学定义，使我们得以把面积算出来(对照

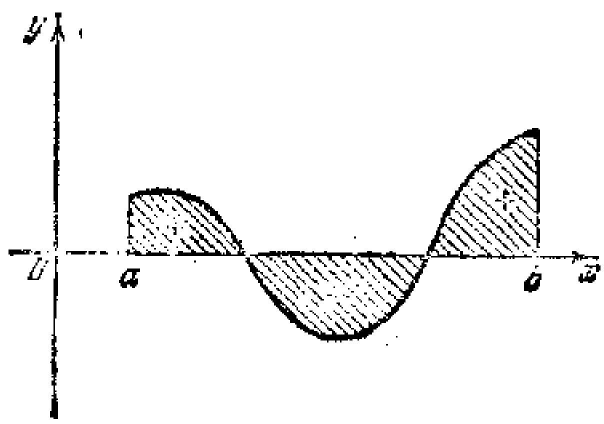


图 25

第 104 页上关于圆周长度及第 109 页上关于速度的话）.

我们在上面假定 $f(x)\geqslant 0$. 若 $f(x)$ 改变正负号，如图 25，则极限(29)所给出的是位于曲线 $y=f(x)$ 与 x 轴之间的各块面积的代数和，其中位于 x 轴之上的各块面积都当作带有正号，而位于 x 轴之下的带有负号.

定积分 还有其他许多问题也使得极限(29)的计算成为必要. 例如，设点在直线上以变速 $v=f(t)$ 运动. 怎样来确定从 $t=a$ 到 $t=b$ 一段时间内点所经历的路程 s 呢?

我们要假定函数 $f(t)$ 是连续的，也就是说，在小段的时间中速度的变化也小. 让我们把 $[a, b]$ 分成长度为 $\Delta t_1, \Delta t_2, \cdots, \Delta t_n$ 的 n 个小段. 为了要近似地计算在每一小段时间 Δt_i 内所经历的路程，我们把这段时间中的速度当作常量，而且就等于这段时间中任意选定的某一瞬时 ξ_i 的速度. 于是点所经历的全部路程就近似地由和式

$$s_n=\sum_{i=1}^{n} f(\xi_i)\Delta t_i$$

表出，而从 a 到 b 整段时间内所经历路程 s 的准确值我们可以看到就是这样的和在小段愈分愈细时的极限，——这是一个跟(29)同一形式的极限：

$$s=\lim_{\max \Delta t_i\to 0}\sum_{i=1}^{n} f(\xi_i)\Delta t_i.$$

我们可以引述许多具体的问题，它们的解决都归结于类似极限的计算. 我们以后还要碰到这样的问题，但上面的例子已经足够说明这种极限的重要性. 极限(29)称为函数 $f(x)$ 在区间 $[a, b]$ 上所取的**定积分**，记作

$$\int_a^b f(x)\,dx.$$

表达式 $f(x)dx$ 称为被积表达式，a 与 b 为积分限；a 是下限，b 是上限.

微分学与积分学的联系 §1 的例 2 可以作为直接计算定积分的一个例子. 现在我们可以说，那里所考虑的问题归结于计算

定积分

$$\int_a^b ax\,dx.$$

在§3中我们考虑过另外一个例子，从而求得了由抛物线 $y=x^2$ 所围成的面积，这里，问题归结于计算积分

$$\int_0^1 x^2\,dx.$$

这两个积分所以都能算出，是由于我们已经知道了最初 n 个自然数之和及这些数平方之和的简单公式. 但决不是对每一个函数 $f(x)$ 都能求出(即用简单公式表出)和式(28)，其中点 ξ_i 与增量 Δx_i 是由这种或那种规律给出的，况且在可能这样去求和的场合，它的实现并不是靠了一种普遍的方法，而是靠了各种各样极特殊的、好像是每一个问题所各别地固有的方法.

因此，发生了定积分普遍计算方法的寻求问题. 在历史上，这问题有很长一个时期以具体课题的形式摆在数学家的面前，这种课题就在于找出一种普遍的方法来确定曲线形面积、曲面所围成物体的体积等等.

我们在前面提到过阿基米德已能计算抛物线弓形及其他一些图形的面积. 后来，那些已经知道解法的关于计算面积、体积、物体重心等等同类的个别问题，数目逐渐增多. 但这些问题的普遍解法的创造过程开始时是极缓慢的. 普遍的方法只有在积累了充分多的理论与计算资料之后才能够创造出来，而资料的产生又是跟实践的需要有紧密联系的. 累积与推广的过程在中世纪晚期才加速发展起来，这是直接由于欧洲的旧(封建制度)生产关系崩溃而新(资本主义)生产关系崛起的时代生产力得到蓬勃发展的结果.

有关定积分计算问题的各种事实的积累是跟相应的有关函数求导数问题的研究同时发生的. 读者在§1中已经知道，这一项巨大的准备工作是在十七世纪在牛顿和莱布尼茨的著作中得到完成的. 在这样的意义上我们说牛顿和莱布尼茨是微积分学的创始者.

牛顿和莱布尼茨的主要功绩之一就是在他们的著作中完全阐

明了存在于微分学与积分学之间的深奥联系．特别是这个联系给极广大的一类函数提供了定积分的一般计算方法．

为了要说明这个联系，让我们回到力学的例子上去．

我们假定有一质点沿着直线运动，其速度 $v=f(t)$ 是时间 t 的函数．我们已经知道质点在 $t=t_1$ 与 $t=t_2$ 一段时间内所经历的路程 σ 等于定积分：

$$\sigma=\int_{t_1}^{t_2} f(t)\,dt.$$

此外，我们再假定质点的运动规律为已知，就是把路程 s 与时间 t 的相倚性表达出来的函数 $s=F(t)$ 为已知，其中路程 s 是从直线上所挑定的某一个起点 A 算起的．在 $[t_1, t_2]$ 一段时间内所经历的路程 σ 显然就等于下面的差：

$$\sigma=F(t_2)-F(t_1).$$

这样，从物理的推论我们达到了等式

$$\int_{t_1}^{t_2} f(t)\,dt=F(t_2)-F(t_1),$$

这等式表达了质点的运动规律与速度之间的关系．

由数学的观点看来，像我们从 §5 中所知道的那样，函数 $F(t)$ 可以被定义为这样一个函数，就是它的导数对于考虑中的一切 t 都等于 $f(t)$，即

$$F'(t)=f(t), \quad (a<x<b).$$

这样一个函数称为 $f(t)$ 的一个**原函数**．

应当注意，函数 $f(t)$ 若至少有一个原函数，就同时有无穷多个原函数，因为若 $F(t)$ 是 $f(t)$ 的原函数，则 $F(t)+C$，其中 C 为任意常量，也是 $f(t)$ 的原函数．但这已经把 $f(t)$ 的一切原函数包罗无遗，因为若 $F_1(t)$ 与 $F_2(t)$ 是同一个函数 $f(t)$ 的原函数，则两者的差 $\varphi(x)=F_1(t)-F_2(t)$ 在所考虑的 t 的变化区间上处处有导数 $\varphi'(t)$ 等于零，因此是一个常量[1]．

---

1) 据中值定理，得

$$\varphi(t)-\varphi(t_0)=\varphi'(\nu)(t-t_0)=0,$$

其中 ν 在 t 与 t_0 之间．据此，对一切 t 有 $\varphi(t)=\varphi(t_0)=$ 常数．

由物理的观点看来，因常量 C 取不同的值而确定的各种运动规律，其差别只在于起算路程的起点 O 也随着而有各种不同的地位.

从以上所说，可以得到这样的结论：设 $f(x)$ 已在区间 $[a, b]$ 上给定并且满足一些极普遍的条件，只要函数 $f(x)$ 在这些条件下可以当作是质点在瞬时 x 的运动速度，就有下列等式成立：

$$\int_a^b f(x)dx = F(b) - F(a), \tag{30}$$

式中 $F(x)$ 是 $f(x)$ 的任何一个原函数[1].

这等式就是著名的**牛顿与莱布尼茨公式**，这公式把计算一个函数的定积分问题归结为寻求该函数的原函数问题，因此在一式之中就使微分学与积分学发生了联系.

曾经成为大数学家研究对象的许多特殊问题，借助于这个公式就自动地得到解决，而这公式所表达的不过是函数 $f(x)$ 在区间 $[a, b]$ 上的定积分等于这一函数的任何一个原函数在区间左右两端上的数值之差[2]. 差式(30)通常还写成下面的样子：

$$F(x)\Big|_a^b = F(b) - F(a).$$

例 1 等式

$$\left(\frac{x^3}{3}\right)' = x^2$$

表示函数 $\frac{x^3}{3}$ 是函数 x^2 的原函数. 因此，根据牛顿与莱布尼茨公式，

$$\int_0^a x^2 dx = \frac{x^3}{3}\Big|_0^a = \frac{a^3}{3} - \frac{0}{3} = \frac{a^3}{3}.$$

例 2 设 e 与 e' 是在一条直线上的两个电荷，彼此相隔的距离为 r. 沿着这条直线在两者之间互相作用的力 F 等于

---

1) 我们可以不求助于力学的例子而用数学方法来证明：若函数 $f(x)$ 在 $[a, b]$ 上连续（或甚至于虽间断但按勒贝格的理论可积，见第三卷第十五章），则必有 $f(x)$ 的原函数 $F(x)$ 存在以及等式(30)成立.

2) 这公式已得到了各种不同的推广（例如，见 § 13，奥斯特洛格拉特斯基公式）.

$$F=\frac{a}{r^2}$$

($a=kcc'$, 式中 k 是一个常量). 当电荷 c 静止不动而电荷 c' 在区间 $[R_1, R_2]$ 上移动时, 我们把区间 $[R_1, R_2]$ 分作小段 Δr_i, 就可以来计算这个力所作的功 W 了. 在每一个小段上我们近似地把力当作常量, 于是在这样的小段上的功等于 $\frac{a}{r_i^2}\Delta r_i$. 使小段愈分愈细, 我们便确信功 W 等于下列积分:

$$W=\lim_{n\to\infty}\sum_{i=1}^{n}\frac{a}{r_i^2}\,\Delta r_i=\int_{R_1}^{R_2}\frac{a}{r^2}\,dr.$$

这个积分我们立刻可以求出, 只要注意到

$$\frac{a}{r^2}=\left(-\frac{a}{r}\right)',$$

就得到

$$W=-\frac{a}{r}\Big|_{R_1}^{R_2}=a\left(\frac{1}{R_1}-\frac{1}{R_2}\right).$$

特别是, 当起先与电荷 c 相距 R_1 的电荷 c' 被移到了无穷远处的时候, 力 F 所作的功等于

$$W=\lim_{R_2\to\infty}a\left(\frac{1}{R_1}-\frac{1}{R_2}\right)=\frac{a}{R_1}.$$

从我们导出牛顿与莱布尼茨公式的推论中, 就显然可见这公式是在数学上表达了客观现实中的一种确定而深奥的联系. 牛顿与莱布尼茨公式正是数学本身反映出客观规律性的一个美妙而且极重要的例子.

应该说明, 牛顿是站在物理的观点上来进行他的数学研究的. 他的创立微积分学原理的著作是跟他的创立力学原理的著作不可分割的.

像导数、积分那样的数学分析概念本身在牛顿和他的同时代人的观念中还完全“不曾脱离”它们的物理与几何原型(速度、面积). 它们在实际上带有一半数学而一半物理的性质. 问题在于当时对于这些概念所下的定义由数学的观点看来还是不能使人满意的, 所以对这些概念的正确运用在稍许复杂的场合就需要研究者有本领甚至在中间的推理阶段上不脱离问题的具体的一面.

由这个观点看来，牛顿与莱布尼茨的创作的性质是不相同的[1]，牛顿在其研究的所有阶段上始终受着物理观点的指导．莱布尼茨的研究却跟物理学没有这样密切的直接关系，这在缺乏清晰数学定义的情况下有时候在个别的研究阶段上引导他到错误的结论．另一方面，莱布尼茨的创作所独有的特征在于力图普遍化，力图找到尽可能更普遍的方法来解决数学分析的问题．

莱布尼茨的最重大功绩在于创造了反映出事物本质的数学符号．数学分析的那些基本概念的记法，例如微分 dx，二阶微分 d^2x，积分 $\int y\,dx$，导数 $\frac{d}{dx}$ 都是莱布尼茨所提出的．这些记号至今还在沿用，足见它们是多么的确切和方便．

妥善地选定的符号可以大大地促使我们的计算与推论变得迅速而容易．它甚至有时候还可以使我们免于堕入错误的结论．对此有充分理解的莱布尼茨在他的创作中就非常注意记号的选择．

数学分析的概念(导数、积分等等)当然在牛顿和莱布尼茨之后继续发展而且直到我们的时代还在发展．但是我们要着重指出在这发展过程中的一个确定的阶段，这一阶段发生在上一世纪的初叶并且首先是跟柯西的著作有关系的．

柯西提供了极限概念的清晰、形式的定义而且在这基础上提供了连续性、导数、微分与积分概念的定义．这些定义都已经在这一章的适当的地方一一引述．在近世分析中我们广泛地使用着这些定义．

这些成就的重要性在于，这样一来，不仅在算术、代数与初等几何中，而且还在一个新的极广阔的数学领域，即数学分析中，也可能纯粹形式地进行运算而同时得到正确的结果．

关于数学分析的一批主要结果在任何场合上的应用，现在我们可以这样说：若起始的数据实际上正确，则数学推论的结果也实际上正确；若我们相信起始的数据充分准确，则所得到结果的正确

1) 牛顿和莱布尼茨的发明是各自独立地得到的．

性不必由实践来验证，而只要检验形式推论的正确性就足够了.

以上所说自然还需要下面的保留，在数学推论中，我们从实践取来的起始数据只在某些误差的限度内才是正确的. 这就引导到如下的情况：当我们对实际数据进行数学的推论时，在每一步上所得到的结果又带有某些误差，同时这种误差随着推论步数的增多而积累起来[1].

回到定积分，让我们来看一看下面这个具有根本重要性的问题. 对于怎样一些在区间$[a, b]$上给定的函数$f(x)$，我们可以保证有定积分$\int_a^b f(x)dx$存在，也就是可以保证有这样一个数存在，当$\max \Delta x_i \to 0$时和式$\sum_1^n f(\xi_i)\Delta x_i$就趋于这个数？我们要注意，这一个数，不论区间$[a, b]$怎样分法，也不论点$\xi_i$怎样取法，都应当是相同的.

那些函数，它的定积分也就是极限(29)确实存在的，称为在区间$[a, b]$上可积. 上一世纪的相应研究已经显示所有的连续函数都是可积的.

也有间断而可积的函数. 例如，在区间$[a, b]$上单调而有界的函数就属于这种函数之列.

在$[a, b]$的有理点上等于0、无理点上等于1的函数可以作为不可积函数的一个例子，因为不论把区间怎样划分，积分和s_n将等于0或1，但看我们取作ξ_i的是有理数还是无理数.

我们要指出，关于怎样实地求出定积分的问题，在许多场合牛顿-莱布尼茨公式已提供了解答. 但这里产生了另外一个问题，就是怎样从已给的函数求出原函数，即以已给函数作为导数的函数. 我们就要转到这一个问题上去. 顺便可以指出，在其他的数学问题中，特别是在求解微分方程时原函数的寻求也具有重大的意义.

1) 例如，由公式$a=b$及$b=c$可以推知$a=c$. 在实践中这一关系是这样的：由$a=b$正确到ε及$b=c$正确到ε，可以推知$a=c$正确到2ε.

§11. 不定积分. 积分的技术

在数学中我们通常把已给函数$f(x)$的任意原函数称为$f(x)$的不定积分，并记作

$$\int f(x)dx.$$

因此，若$F(x)$是$f(x)$的某一个完全确定的原函数，则$f(x)$的不定积分就等于

$$\int f(x)dx=F(x)+C, \tag{31}$$

式中C是任意常数.

我们还要指出，若函数$f(x)$在区间$[a, b]$上给定，$F(x)$是它的原函数而x是区间$[a, b]$的点，则根据牛顿-莱布尼茨公式，可以写出

$$F(x)=F(a)+\int_a^x f(t)dt.$$

这样，在等式右面的积分跟$f(x)$的原函数$F(x)$只相差一个常量$F(a)$. 在这样的场合，这一个积分，(在x可变时)当作上限x的函数看待，正是$f(x)$的某一个完全确定的原函数，因而$f(x)$的不定积分还可以写成下面的形式：

$$\int f(x)dx=\int_a^x f(t)dt+C,$$

式中C是任意常量.

让我们把基本的不定积分表列举于下，这是直接从相应的导数表(见§6)编成的：

$$\int x^a dx=\frac{x^{a+1}}{a+1}+C \quad (a\neq -1), \qquad \int \cos x\, dx=\sin x+C,$$

$$\int \sec^2 x\, dx=\operatorname{tg} x+C, \qquad \int a^x dx=\frac{a^x}{\ln a}+C,$$

$$\int\frac{dx}{x}=\ln|x|+C^{1)},\qquad \int e^x\,dx=e^x+C,$$

$$\int\frac{dx}{\sqrt{1-x^2}}=\arcsin x+C=-\arccos x+C_1\quad\left(C_1-C=\frac{\pi}{2}\right),$$

$$\int\sin x\,dx=-\cos x+C,\qquad \int\frac{dx}{1+x^2}=\operatorname{arc\,tg} x+C.\tag{32}$$

不定积分的一般性质也可以根据导数的相应的性质推出来。例如，由和的微分法则我们得到公式

$$\int[f(x)\pm\varphi(x)]\,dx=\int f(x)\,dx\pm\int\varphi(x)\,dx+C,$$

而由常数因子 k 可从导数符号下移出的法则得到

$$\int kf(x)\,dx=k\int f(x)\,dx+C.$$

这样

$$\int\left(3x^2+2x-\frac{3}{\sqrt{x}}+\frac{4}{x}-1\right)dx$$

$$=3\frac{x^3}{3}+\frac{2x^2}{2}-3\frac{x^{-\frac{1}{2}+1}}{-\frac{1}{2}+1}+4\ln|x|-x+C$$

$$=x^3+x^2-x-\frac{2}{3}\sqrt{x}+\ln x^4+C.$$

我们有好些计算不定积分的方法。让我们来看一看其中的一种方法，就是变量置换法或换元法[2)]，这是以下列等式的正确性作为根据的：

$$\int f(x)\,dx=\int f[\varphi(t)]\varphi'(t)\,dt+C,\tag{33}$$

式中 $x=\varphi(t)$ 是可微函数。关系式(33)应当这样加以理解：若在跟(33)式左面相等的函数

$$F(x)=\int f(x)\,dx$$

---

1) 当 $x>0$ 时，$\qquad (\ln|x|)'=(\ln x)'=\frac{1}{x}$，

当 $x<0$ 时，$\qquad (\ln|x|)'=[\ln(-x)]'=\frac{1}{-x}(-1)=\frac{1}{x}$.

2) 变量又称变元，或简称为元。——译者注

中令 $x=\varphi(t)$，则得到这样一个函数 $F[\varphi(t)]$，它对 t 的导数就等于(33)式右面积分符号下的表达式. 这是可以从关于复合函数导数的定理直接推出来的.

让我们就换元法的应用再举一些例子：

$$\int e^{kx}\,dx=\int e^t\frac{1}{k}\,dt=\frac{1}{k}\int e^t dt=\frac{1}{k}e^t+C=\frac{e^{kx}}{k}+C$$

(令 $kx=t$，从而有 $k\,dx=dt$).

$$\int\frac{x\,dx}{\sqrt{a^2-x^2}}=-\int dt=-t+C=-\sqrt{a^2-x^2}+C$$

$\left(令\ t=\sqrt{a^2-x^2}，从而有\ dt=-\frac{x\,dx}{\sqrt{a^2-x^2}}\right)$.

$$\begin{aligned}\int\sqrt{a^2-x^2}\,dx&=\int\sqrt{a^2-a^2\sin^2u}\,a\cos u\,du\\&=a^2\int\cos^2u\,du=a^2\int\frac{1+\cos 2u}{2}\,du\\&=\frac{a^2}{2}\left(u+\frac{\sin 2u}{2}\right)+C\\&=\frac{a^2}{2}(u+\sin u\cos u)+C\\&=\frac{a^2}{2}\left(\arcsin\frac{x}{a}+\frac{x}{a^2}\sqrt{a^2-x^2}\right)+C\end{aligned}$$

(令 $x=a\sin u$).

从这些例子显然可见，换元法大大地扩充了我们现在所能够加以积分的那种初等函数，所谓能够加以积分就是说能够求得仍属初等函数的原函数. 但应当注意，由计算的观点看来，一般地说，积分的处境远比微分来得恶劣.

由§6已经知道任何初等函数的导数又是初等函数，这是利用微分法则就可以完全有效地求得到的. 但是相反的论断，一般地说，并不真确，因为存在着那样的初等函数，它的不定积分不再是初等函数. 例如，e^{-x^2}，$\frac{1}{\ln x}$，$\frac{\sin x}{x}$ 等就是这样的函数. 要得到这些函数的积分，不得不使用近似方法，并且引入一些无法化成初等函数的新函数. 我们不可能在这个问题上逗留很久，我们只想

指出，在初等数学中就可以找到许多例子，其中一种运算可以在某一类的数中施行而它的逆运算就不能在同一类中实现；如任何正有理数的平方是有理数，但有理数的平方根就决非始终是有理数．类似地，对初等函数微分，所得到的还是初等函数，但对初等函数积分就可以使我们脱离这一类函数了．

有些不能用初等函数来计算的积分在数学及其应用中具有重大的意义．积分

$$\int_0^x e^{-t^2}dt$$

就是这样的一个例子，它在概率论中起着极巨大的作用(见第二卷第十一章)．我们还要指出积分

$$\int_0^\varphi \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} \quad 与 \quad \int_0^\varphi \sqrt{1-k^2\sin^2\theta}\,d\theta \quad (k^2<1)$$

分别有第一与第二种**椭圆积分**之称．有极多的力学与物理学问题归结为这两个积分的计算(见第二卷第五章§1，例3)．就自变量 k 与 φ 取各种不同的值，已经编造出这些积分值的明细表，这是用近似方法算出的，但有很高的准确度．

应当强调指出，要证明这个或那个初等函数是无法用初等函数来积分的，在每一个各别的场合引起极大的困难．上一世纪的一些杰出的数学家与分析学家都竭其才智于这些问题，而这些问题的研究也在分析的发展中起了重大的作用．在这方面车比雪夫曾经得到一些主要的结果，特别是他充分研究了如下形式的积分是不是可能用初等函数来积分的问题：

$$\int x^m(a+bx^s)^p\,dx,$$

式中 m, s 与 p 都是有理数．在车比雪夫之前，牛顿早就得到可使这个积分用初等函数积出的指数 m, s 及 p 之间的三种关系，车比雪夫证明了在所有其他的场合这积分都是不能用初等函数表达的．

我们还要引述另一种积分方法——分部积分法．这是以读者所已经知道的关于两函数 u 与 v 之积的导数公式：

$$(uv)'=uv'+u'v$$

作为根据的．这公式还可以这样写：

$$uv'=(uv)'-u'v.$$

现在让我们对左右两边分别积分，并且注意

$$\int (uv)'dx=uv+C,$$

最后就得到等式

$$\int uv'\,dx=uv-\int u'v\,dx,$$

这个公式称为分部积分公式（我们没有写出常数 C，因为可以把它看作是已经包括在等式中的一个不定积分里）．

让我们就这公式的应用举一些例子．试计算 $\int xe^x\,dx$．在这个积分中可令 $u=x$ 而 $v'=e^x$，于是 $u'=1$，$v=e^x$，从而有

$$\int xe^x\,dx=xe^x-\int 1\cdot e^x\,dx=xe^x-e^x+C.$$

在积分 $\int \ln x\,dx$ 中，令 $u=\ln x$，$v'=1$ 方才适当，于是 $u'=\frac{1}{x}$，$v=x$，而

$$\int \ln x\,dx=x\ln x-\int dx=x\ln x-x+C.$$

还有一个特异的例子，须两次分部积分，然后从所得到的等式中获得所求的积分：

$$\int e^x\sin x\,dx=e^x\sin x-\int e^x\cos x\,dx$$

$$=e^x\sin x-e^x\cos x-\int e^x\sin x\,dx,$$

从而有 $$\int e^x\sin x\,dx=\frac{e^x}{2}(\sin x-\cos x)+C.$$

我们就以这个例子作为这一节的结尾，读者从这一节中仅仅得到了积分理论的一些肤浅的观念．这理论的许多方法我们都不曾讲到，特别是我们在这里不曾接触到极饶趣味的有理分式积分理论，上一世纪的著名数学家与力学家奥斯特洛格拉特斯基曾对该理论作出重大的贡献．

§12. 多元函数

到现在为止我们所讲的都是一元函数，但在实践中我们往往还必须碰到由二元、三元以及一般地说由多元来确定的函数．例如，矩形的面积是其底 x 与高 y 的函数：

$$S=xy.$$

长方体的体积是其长、宽、高的函数：

$$v=xyz.$$

两点 A 与 B 之间的距离是这两点的六个坐标的函数：

$$r=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2}.$$

大家所知道的公式

$$pv=RT$$

表达了一定数量的气体的体积 v 怎样随着它的压力 p 与绝对温度 T 而定.

正像一元函数一样，多元函数常常只在变量在某一区域上取值时才给出．例如，函数

$$u=\ln(1-x^2-y^2-z^2) \tag{34}$$

仅在 x, y, z 的值满足条件

$$x^2+y^2+z^2<1 \tag{35}$$

时方才给定(对于其他的 x, y, z，函数值不是实数)．坐标满足不等式(35)的空间点集显然填满了一个半径为 1 而中心在原点的球．边界的点不属于这个球．它的表面仿佛"已被剥掉"．这样的球称为开球．函数(34)只在(x, y, z)三个数是这个开球 G 各点的坐标时才有定义．通常我们就简短地说，函数(34)定义于开球 G.

这里是另外一个例子．一个已受到不均匀加热的物体 V 的温度是这物体上各点的坐标 x, y, z 的某一个函数．这函数并未对所有(x, y, z)三个数，而只在(x, y, z)是物体 V 各点的坐标时才有定义.

最后，作为第三个例子，让我们考虑函数

$$u=\varphi(x)+\varphi(y)+\varphi(z),$$

式中 φ 是定义于区间 $[0, 1]$ 的一元函数. 显然函数 u 只在 (x, y, z) 三个数是立方体:

$$0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 1,\ 0\leqslant z\leqslant 1$$

各点的坐标时才有定义.

让我们把三元函数的定义提出来. 设已知 (x, y, z) 三个数的集(空间点集)E. 若对应于 E 的每三个数(每点), 根据某一法则有一个确定的数 u, 则我们说 u 是 x, y, z(点)的函数, 定义于三个数的集(点集)E, 记作

$$u=F(x, y, z).$$

除了 F 外, 也可以用其他的字母: f, φ, ψ.

在实践中作为集 E 的往往是填满某些几何体(区域), 如球、立方体、环等等的点集, 于是我们就简单地说函数定义于这个几何体(区域). 我们可以类似地来定义二元、四元等等函数.

隐函数 我们要指出, 二元函数在某种情况下可以很好地作为给出一元函数的工具. 设已知二元函数 $F(x, y)$, 让我们构成方程

$$F(x, y)=0. \tag{36}$$

一般地说, 它确定了 (x, y) 平面的某一个 (x, y) 点集, 在这些点上我们的函数都等于零. 往往这样的点集就是某一条曲线, 这曲线可以当作是一个或几个单值一元函数 $y=\varphi(x)$ 或 $x=\psi(y)$ 的图形. 在这样的场合, 我们说, 这些单值的函数是借助于方程(36)而被隐含地确定的[1]. 例如, 方程

$$x^2+y^2-r^2=0.$$

就以隐含的方式确定了两个一元函数:

$$y=+\sqrt{r^2-x^2} \quad 与 \quad y=-\sqrt{r^2-x^2}.$$

但是应当注意, 像(36)那样的方程也可以不确定任何函数. 例如, 方程

1) 我们也往往这样说: 在方程 $F(x, y)=0$ 中 y (或 x) 是 x (或 y) 的隐函数. ——译者注

$$x^2+y^2+1=0$$

显然并未给出任何实函数，因为没有一对实数能够满足这个方程．

几何的表示 二元函数可以借助于空间坐标系而很直观地由曲面来表示．这样，函数

$$z=f(x, y) \quad (37)$$

就在三维空间直角坐标系中被绘成一个曲面，这曲面是坐标满足方程(37)的点 M 的轨迹(图 26)．

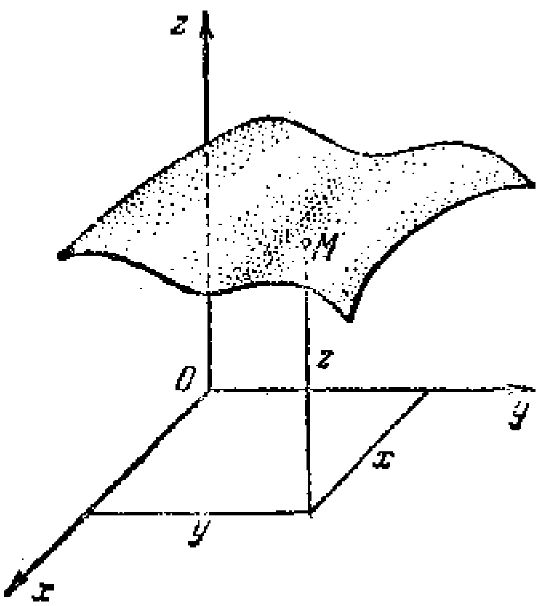


图 26

另外还有一种极方便的表示函数(37)的方法，在实践中早已得到广泛的应用．取定了一连串的值 $z_1, z_2, \cdots$，我们就在同一个 Oxy 平面上绘出曲线：

$$z_1=f(x, y), \quad z_2=f(x, y), \cdots,$$

就是所谓函数 $f(x, y)$ 的等高线．按照等高线，只要它们相应于彼此充分接近的 z 值，就可以很好地看出函数 $f(x, y)$ 的变化，正如按照地形图上的等高线就可以把地势的变化测定．

在图 27 中已经绘出了函数 $z=x^2+y^2$ 的等高线图；在旁边又

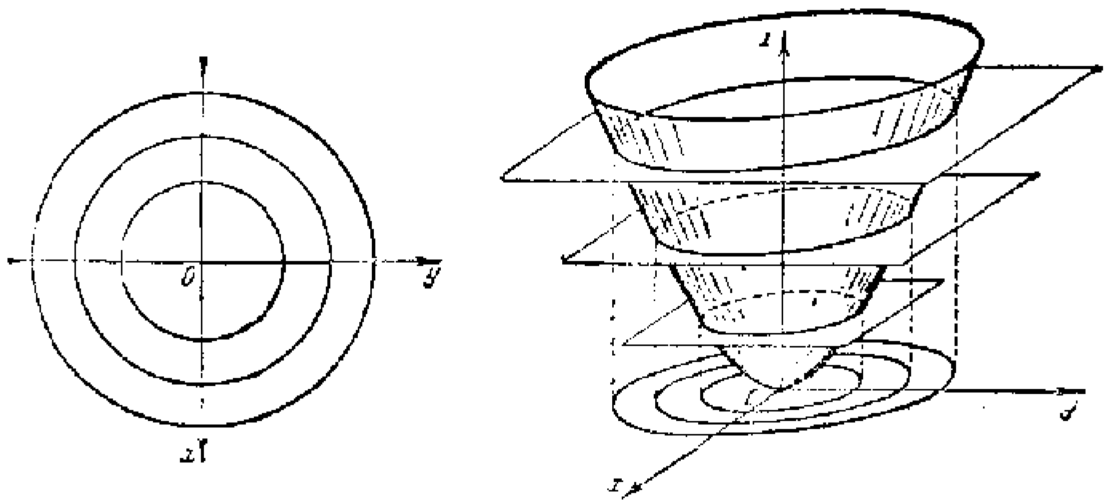


图 27

显示了这些线是怎样作出的. 在第三章图 50 (第 232 页) 中给函数 $z=xy$ 绘出了同样的等高线图.

偏导数与微分 让我们简略地说一说多元函数的微分法. 我们以任意的二元函数

$$z=f(x, y)$$

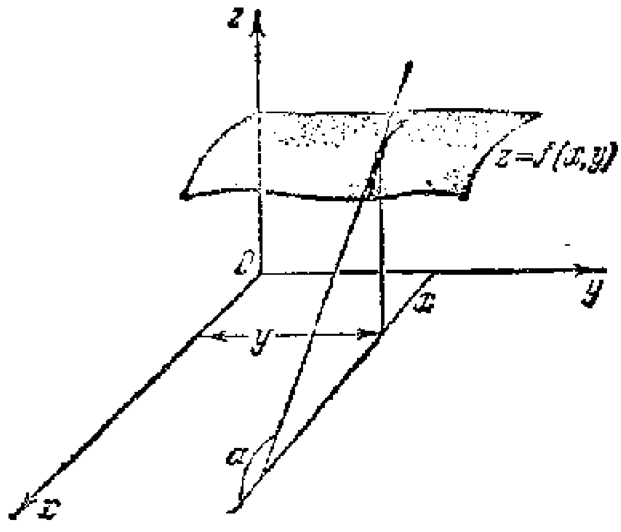


图 28

作为例子. 若把 y 的值固定, 就是说, 把 y 当作是不变的, 则我们的二元函数就变为一元 x 的函数. 它的导数, 如果存在, 称为对 x 的偏导数, 记作

$$\frac{\partial z}{\partial x} \text{ 或 } \frac{\partial f}{\partial x} \text{ 或 } f'_x(x, y).$$

最后一种记法强调指出, 对 x 的偏导数, 一般地说, 是 x 和 y 的函数. 我们可以类似地来定义对 y 的偏导数.

在几何上我们的函数是由空间直角坐标系里的一个曲面来表示的. 把 y 固定后所得到的 x 的函数则由一条平面曲线表出, 这曲线是由那个与 Oxz 平面平行而且相距 y 的平面跟曲面相截而成 (图 28). 偏导数 $\frac{\partial z}{\partial x}$ 显然就等于这曲线在 (x, y) 点上的切线跟 x 轴正向所成交角的正切.

一般地说, 若已给 n 元 $x_1, \cdots, x_n$ 的函数 $z=f(x_1, x_2, \cdots, x_n)$, 则在其他各元:

$$x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n$$

的值作为固定时对 x_i 所算出的导数称为偏导数 $\frac{\partial z}{\partial x_i}$.

我们可以说, 一个函数对变元 x_i 的偏导数是这函数在 x_i 变化方向的变化率. 我们还可以来定义对任意给定方向的导数, 所谓任意给定的方向就不必一定跟这个或那个坐标轴相同, 但我们不拟加以细说.

例　1) $z=\frac{x}{y}$, $\frac{\partial z}{\partial x}=\frac{1}{y}$, $\frac{\partial z}{\partial y}=-\frac{x}{y^2}$.

2) $u=\frac{1}{\sqrt{x^2+y^2+z^2}}$,

$$\frac{\partial u}{\partial x}=-\frac{1}{x^2+y^2+z^2}\frac{2x}{2\sqrt{x^2+y^2+z^2}}=-\frac{x}{(x^2+y^2+z^2)^{3/2}}.$$

有时候我们还需要对偏导数再求偏导数——这就是所谓二阶偏导数. 二元函数有四个这样的偏导数:

$$\frac{\partial^2 u}{\partial x^2},\quad \frac{\partial^2 u}{\partial x\,\partial y},\quad \frac{\partial^2 u}{\partial y\,\partial x},\quad \frac{\partial^2 u}{\partial y^2}.$$

但是在这些导数连续的场合,上面所写的第二和第三个(所谓混合导数)可以证明是相同的:

$$\frac{\partial^2 u}{\partial x\,\partial y}=\frac{\partial^2 u}{\partial y\,\partial x}.$$

例如,就前面所考虑的第一个函数而论,

$$\frac{\partial^2 z}{\partial x^2}=0,\quad \frac{\partial^2 z}{\partial x\,\partial y}=-\frac{1}{y^2},\quad \frac{\partial^2 z}{\partial y\,\partial x}=-\frac{1}{y^2},\quad \frac{\partial^2 z}{\partial y^2}=\frac{2x}{y^3},$$

读者可立即看到混合导数是相同的.

对于多元函数,正像对于一元函数一样,我们也可以引入微分概念.

为明确起见,让我们来看二元函数

$$z=f(x,\ y).$$

若它有连续的偏导数,则我们可以证明,相应于自变量增量 Δx 与 Δy 的函数增量

$$\Delta z=f(x+\Delta x,\ y+\Delta y)-f(x,\ y)$$

可以写成　$\Delta z=\frac{\partial f}{\partial x}\Delta x+\frac{\partial f}{\partial y}\Delta y+\alpha\sqrt{(\Delta x)^2+(\Delta y)^2}$,

其中 $\frac{\partial f}{\partial x}$ 与 $\frac{\partial f}{\partial y}$ 是函数在点$(x,\ y)$上的偏导数,而量 α 随着 Δx 与 Δy 而定,并且当 $\Delta x\to 0$, $\Delta y\to 0$ 时 $\alpha\to 0$.

上式右面最初两项之和

$$dz=\frac{\partial f}{\partial x}\Delta x+\frac{\partial f}{\partial y}\Delta y$$

线性地[1]随着 Δx 与 Δy 而定，称为函数的微分．最后一项里边有着跟 Δx 与 Δy 一起趋零的因子 α，所以就表明 x 与 y 的共同变化的量

$$\rho=\sqrt{(\Delta x)^2+(\Delta y)^2}$$

而论，这一项是关于 ρ 的高阶无穷小．

这里是微分概念应用的一个例子．钟摆振动的周期是按照下列公式来计算的：

$$T=2\pi\sqrt{\frac{l}{g}},$$

其中 l 是摆的长度而 g 是重力加速度．我们假定已知 l 与 g 分别有误差 Δl 与 Δg．于是我们在计算 T 时所招致的误差将等于与自变量增量 Δl 和 Δg 相对应的增量 ΔT．近似地以 dT 替代 ΔT，则有

$$\Delta T\approx dT=\pi\left(\frac{\Delta l}{\sqrt{lg}}-\frac{\sqrt{l}\,\Delta g}{\sqrt{g^3}}\right).$$

我们不知道 Δl 与 Δg 的正负号，但显然可以从下面的不等式来估计 ΔT：

$$|\Delta T|\leqslant\pi\left(\frac{|\Delta l|}{\sqrt{lg}}+\sqrt{\frac{l}{g^3}}\,|\Delta g|\right),$$

而在两面各除以 T 后，得

$$\frac{|\Delta T|}{T}\leqslant\frac{1}{2}\left(\frac{|\Delta l|}{l}+\frac{|\Delta g|}{g}\right).$$

所以，在实际上可以认为 T 的相对误差就等于 l 与 g 的相对误差的平均值．

为了记法的对称起见，自变量的增量 Δx 与 Δy 通常记作 dx 与 dy，并且也称为这些自变量的微分．用了这样的记法，函数 $u=f(x, y, z)$ 的微分就可写成：

$$du=\frac{\partial f}{\partial x}dx+\frac{\partial f}{\partial y}dy+\frac{\partial f}{\partial z}dz.$$

每次当我们必须跟多元函数发生关系的时候（而这在分析对

1）一般地说，函数 $Ax+By+C$（其中 A, B, C 为常数）称为 x 与 y 的线性函数．若 $C=0$，则称为齐次线性函数．在这里我们省略了“齐次”字样．

技术与物理学问题的大量应用中是常有的事)，偏导数起着巨大的作用．关于怎样从一个函数的偏导数的性质来恢复这个函数的问题，我们还将在第二卷第六章中看到．

下面我们就偏导数在分析中的应用举出一些最简单的范例．

隐函数求导法 设有 x 的隐函数 y 由方程

$$F(x,\ y)=0 \tag{38}$$

给出，而我们需要知道这函数的导数．若 $x,\ y$ 满足等式(38)而我们给 x 以增量 Δx，则 y 将得到那样的增量 Δy，使得 $x+\Delta x$ 与 $y+\Delta y$ 又满足(38)．所以[1)]

$$\begin{aligned}&F(x+\Delta x,\ y+\Delta y)-F(x,\ y)\\&\quad=\frac{\partial F}{\partial x}\Delta x+\frac{\partial F}{\partial y}\Delta y+\alpha\sqrt{(\Delta x)^2+(\Delta y)^2}=0.\end{aligned}$$

据此，只要 $\dfrac{\partial F}{\partial y}\neq 0$，就有

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=y'_x=-\frac{\dfrac{\partial F}{\partial x}}{\dfrac{\partial F}{\partial y}}.$$

这样，我们得到了一个对 x 的隐函数 y 求导数的方法，可不必先把方程(38)对 y 解开．

极大与极小问题 若一个函数，比如说二元函数 $z=f(x,\ y)$，在点 $(x_0,\ y_0)$ 达到极大，也就是说，若对于邻近 $(x_0,\ y_0)$ 的一切点 $(x,\ y)$ 有 $f(x_0,\ y_0)\geqslant f(x,\ y)$，则这一点也应当是曲面 $z=f(x,\ y)$ 被平行于 Oxz 或 Oyz 的平面所截出的曲线上的极高点．所以在这样的点上必然满足条件

$$f'_x(x,\ y)=0,\quad f'_y(x,\ y)=0. \tag{39}$$

在局部极小的点上也应当满足同样的条件．所以函数的最大与最小值首先应当在满足条件(39)的点中去寻求．(此外，不要忘记在函数定义域边界上的点以及函数的导数不存在的点，如果有的话．)

---

1) 我们假定 $F(x,\ y)$ 有连续的对 x 与 y 的导数．

为了要确定所找到的满足条件(39)的点实际上是不是极大或极小点，往往要利用各种各样间接的推论．例如，若根据某种理由已知函数可微并且在区域内部达到极小，而在那里只有一个满足条件(39)的点，则显然就在这一点上达到极小．

举一个具体的例子来说，假设我们要用铁皮做一个(无盖的)具有已知容积 V 的长方盒，使所耗费的材料尽可能地少．若把这盒底的两边记作 x 与 y，则盒高 h 将等于 $\frac{V}{xy}$，因而盒的表面积 S 将由 x 与 y 的函数：

$$S=xy+\frac{V}{xy}(2x+2y)=xy+2V\left(\frac{1}{x}+\frac{1}{y}\right) \tag{40}$$

表出．因为 x 与 y 按照题意应当是正的，所以问题归结于从(x,y)坐标平面第一象限内的一切点(x, y)中求出函数 $S(x, y)$的极小，这一个区域我们记作 G．

若在区域 G 的某一点上函数达到极小，则在这一点上偏导数必须等于零：

$$\frac{\partial S}{\partial x}=y-\frac{2V}{x^2}=0, \quad \frac{\partial S}{\partial y}=x-\frac{2V}{y^2}=0,$$

即 $yx^2=2V$，$xy^2=2V$，从而得到盒的长、宽、高：

$$x=y=\sqrt[3]{2V} \quad 及 \quad h=\sqrt[3]{\frac{V}{4}}. \tag{41}$$

我们已经解决了所提出的问题，但是完全不曾阐明解答的根据．严格的数学家将对我们说："你们一开始就假定在所设条件下有表面积最小的盒子存在，而从这个假定出发，找到了它的长、宽、高．因此，直到现在你们所得到的还只是这样的论断：如果在 G 内有使函数 S 达到极小的点(x, y)存在，那末它的坐标必须由等式(41)来确定．请先证明 S 在 G 上有极小存在，我才认为你们的结果正确．"这意见是有充分道理的，因为就拿我们的函数 S 来说，我们马上可以看到它在区域 G 中是没有极大的．我们要指出怎样才可以确认在这一场合我们的函数真正在区域 G 的某一点(x, y)上达到它的极小．

我们将在这里取作根据的而且在分析中可以十分严格地加以证明的基本论断如下所述．若一元或多元函数 f 在某一个由边界围成并且包含边界在内的有限区域 H 上处处连续，则在 H 中总至少有一点，使这函数达到极小（极大）．借助于这个论断，我们就不难彻底地来分析我们的例子．

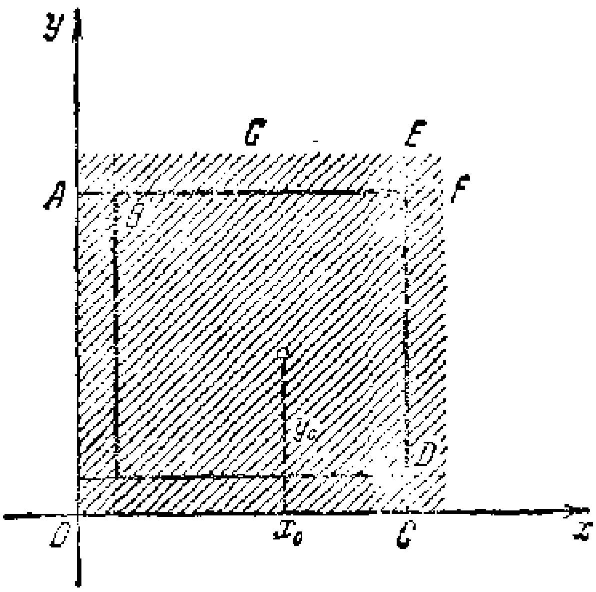


图 29

让我们取定区域 G 的任何一点 (x_0, y_0)，设在这一点上 $S(x_0, y_0)=N$．其次，让我们取定一个同时满足不等式 $R>N$, $2VR>N$ 的数 R，并作一边长为 R^2 的正方形 Ω_R 如图 29，其中 $AB=OD=\frac{1}{R}$．

让我们在区域 G 的位于正方形 Ω_R 外部的点上来估计函数 $S(x, y)$ 的下界．若区域 G 的点有横坐标 $x<\frac{1}{R}$，则

$$S(x, y)=xy+2V\left(\frac{1}{x}+\frac{1}{y}\right)>2V\frac{1}{x}>2VR>N.$$

类似地，若区域 G 的点有纵坐标 $y<\frac{1}{R}$，则也有 $S>N$，其次，若区域 G 的点有横坐标 $x>\frac{1}{R}$ 而位于直线 AF 之上，或者有纵坐标 $y>\frac{1}{R}$ 而位于直线 CE 之右，则

$$S(x, y)>xy>\frac{1}{R}R^2=R>N.$$

这样，对于区域 G 的在正方形 Ω_R 外部的一切点，有不等式 $S(x, y)>N$ 成立，又因为 $S(x_0, y_0)=N$，所以点 (x_0, y_0) 属于正方形，因此我们的函数在 G 上的极小就等于它在正方形上的极小．

但是函数 $S(x, y)$ 在这正方形的内部及边界上连续，于是根据上述论断在正方形中有一点 (x, y)，使我们的函数在正方形上达到极小，因此也就在我们的区域 G 中达到极小．至此，极小的存在已得到证明．

在对定义于无限区域上的函数寻求极大或极小时，我们可以怎样来进行讨论，上面所引用的推论就是一个范例．

泰勒公式 多元函数，也跟一元函数一样，可以用泰勒公式来表达．例如，把函数

$$u=f(x, y)$$

在点 (x_0, y_0) 的邻近展开，若限于 $x-x_0$ 与 $y-y_0$ 的一次及二次幂，就有如下形式：

$$\begin{aligned}f(x, y)=f(x_0, y_0)&+[f'_x(x_0, y_0)(x-x_0)+f'_y(x_0, y_0)(y-y_0)]\\&+\frac{1}{2!}[f''_{xx}(x_0, y_0)(x-x_0)^2+2f''_{xy}(x_0, y_0)(x-x_0)\\&\qquad(y-y_0)+f''_{yy}(x_0, y_0)(y-y_0)^2]+R_3.\end{aligned}$$

这时，若函数 $f(x, y)$ 有连续的二阶偏导数，则当

$$r=\sqrt{(x-x_0)^2+(y-y_0)^2},$$

即点 (x, y) 与 (x_0, y_0) 之间的距离趋零时，余项 R_3 要比 r^2 更迅速地趋于零．泰勒公式提供了一种极普遍的工具，借以给出各种函数并近似地计算它们的数值．

我们要指出，借助于这个公式我们也就可以解决上面所提出的关于在 $\frac{\partial f}{\partial x}=\frac{\partial f}{\partial y}=0$ 的点上函数究竟有没有极大或极小的问题．事实上，若这些条件已在某一点 (x_0, y_0) 上成立，则根据泰勒公式，对于邻近 (x_0, y_0) 的点 (x, y) 来说，函数值将与 $f(x_0, y_0)$ 相差下面的量：

$$\begin{aligned}&f(x, y)-f(x_0, y_0)\\&=\frac{1}{2!}[A(x-x_0)^2+2B(x-x_0)(y-y_0)+C(y-y_0)^2]+R_3.\end{aligned}\tag{42}$$

其中 A，B 与 C 分别记点 (x_0, y_0) 上的二阶偏导数 f''_{xx}，f''_{xy} 与 f''_{yy}．

若我们看到函数

$$\Phi(x, y)=A(x-x_0)^2+2B(x-x_0)(y-y_0)+C(y-y_0)^2$$

对于任意的、不同时等于零的 $x-x_0$ 与 $y-y_0$ 都是正的，则在这种场合等式(42)的整个右面对于微小的 $x-x_0$ 与 $y-y_0$ 将是正的，因为在 $x-x_0$ 与 $y-y_0$ 充分小时，量 R_3 的绝对值显然小于 $\frac{1}{2}\Phi(x, y)$. 由此可见函数 f 这时在点 (x_0, y_0) 达到极小. 反之，若函数 $\Phi(x, y)$ 对于任意的 $x-x_0$ 与 $y-y_0$ 都是负的，则这就使等式(42)的整个右面对于微小的 $x-x_0$ 与 $y-y_0$ 也为负值，因而在点 (x_0, y_0) 上有极大.

在更复杂的场合，必须考虑泰勒公式中二次以后的各项.

关于三元及更多元函数的极大与极小问题也可用完全类似的方法来讨论. 作为一个练习，读者可以自己来证明下面的结果：设在空间已给定的点

$$P_1(x_1, y_1, z_1),\ P_2(x_2, y_2, z_2),\ \cdots,\ P_n(x_n, y_n, z_n)$$

上分别安置着已知的质量

$$m_1,\ m_2,\ \cdots,\ m_n,$$

这一组质量对 $P(x, y, z)$ 点的力矩 M 等于每一质量乘上它到 P 点距离的平方，再一一相加：

$$M(x, y, z)=\sum_{i=1}^{n} m_i[(x-x_i)^2+(y-y_i)^2+(z-z_i)^2],$$

而这一组的所谓重心有坐标

$$x=\frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i},\quad y=\frac{\sum_{i=1}^{n} m_i y_i}{\sum_{i=1}^{n} m_i},\quad z=\frac{\sum_{i=1}^{n} m_i z_i}{\sum_{i=1}^{n} m_i}.$$

若把 P 点放在重心上，M 就取得所有可能值中的极小.

相对极大与极小 对于多元函数还可以提出一种稍微不同的极大与极小问题. 让我们用简单的例子来说明. 设已知一圆，其半径为 R，而我们要从这圆的一切内接矩形中求出一个具有最大面积的矩形. 矩形的面积等于它的两边的乘积 xy，其中 x 与 y 都

是正的而且在这一场合还有关系式 $x^2+y^2=(2R)^2$ 加以联系，如图 30 显然可见. 这样，我们就只要从满足关系式 $x^2+y^2=4R^2$ 的 x 与 y 中去寻求函数 $f(x, y)=xy$ 的极大.

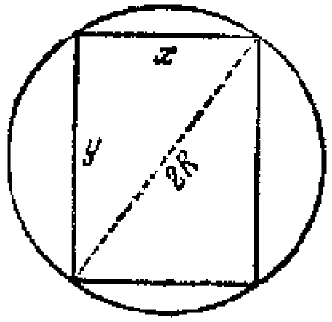


图 30

在实践中很频繁地发生类似的问题，就是我们需要只从满足某一关系式

$$\varphi(x, y)=0$$

的 x 与 y 中求出某一函数 $f(x, y)$ 的极大(或极小).

当然，我们可以先把方程 $\varphi(x, y)=0$ 对 y 解开，把所得到的 y 的表达式代入 $f(x, y)$，然后对一元 x 的函数求通常的极大. 但这一条路往往是复杂的，而且有时候是行不通的.

为解决这一类的问题，在分析中产生了一种更便利的方法，称为拉格朗日乘数法. 它的用意是十分简单的. 让我们考虑函数

$$F(x, y)=f(x, y)+\lambda\varphi(x, y),$$

式中 λ 是一个任意常数. 显然，对于满足条件 $\varphi(x, y)=0$ 的 x 与 y，$F(x, y)$ 的值就跟 $f(x, y)$ 的相同.

对于函数 $F(x, y)$ 我们将在 x, y 没有任何联系的情况下来寻求极大. 在极大点上必须遵守条件 $\frac{\partial F}{\partial x}=\frac{\partial F}{\partial y}=0$[1]，也就是

$$\frac{\partial f}{\partial x}+\lambda\frac{\partial \varphi}{\partial x}=0, \tag{43}$$

$$\frac{\partial f}{\partial y}+\lambda\frac{\partial \varphi}{\partial y}=0. \tag{44}$$

在 $F(x, y)$ 极大点上的 x 与 y 值就是方程组(43), (44)的一组解而且当然跟这些方程中所含的系数 λ 有关系. 现在我们假定已经选定数 λ，使得极大点的坐标满足条件

$$\varphi(x, y)=0. \tag{45}$$

于是这一点也将是原来的问题中的局部极大点.

1) 当然这里所说的是在函数 $F(x, y)$ 的定义域内部所达到的极大，函数 $f(x, y)$ 与 $\varphi(x, y)$ 都假定是可微的.

事实上，我们可以用几何方法像下面这样来看我们的问题。函数$f(x, y)$已在某一个区域G中给定(图 31)。满足条件$\varphi(x, y)=0$的通常是某一条曲线Γ的点。我们要在曲线Γ的点上求出$f(x, y)$的最大值。若$F(x, y)$在曲线Γ的点上达到极大，则从这一点向任何一面作不大的位移，包括沿曲线Γ的位移在内，$F(x, y)$都不增大。但在沿Γ移动时$F(x, y)$的值跟$f(x, y)$的相同，这就是说，在沿曲线作微小位移时函数$f(x, y)$也不增大，因此在这一点上有局部极大。

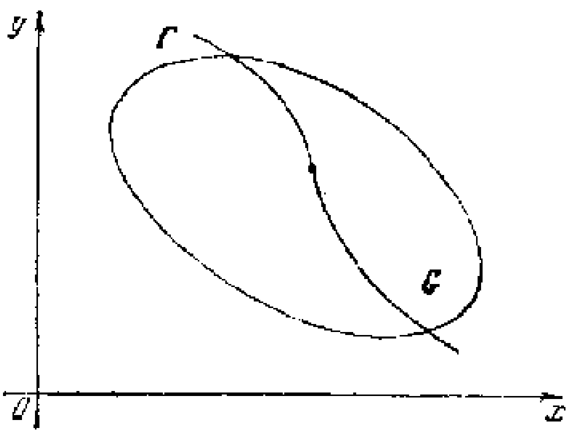


图 31

这些设想提出了解决问题的简单方法，让我们构成方程(43)，(44)，(45)；对未知数x, y, λ来求解这方程组；得到一组或几组解：

$$(x_1, y_1, \lambda_1), (x_2, y_2, \lambda_2), \cdots. \tag{46}$$

在点(x_1, y_1)，(x_2, y_2)，$\cdots$之外，还要加入曲线Γ在穿过区域G时所碰到的G的边界点，然后从所有这些点中挑出那一个使$f(x, y)$取得最大(或最小)值的点。

当然，我们的启发推论还丝毫不曾证明这个方法的正确性。事实上，我们并未证明，当λ取某值时所得到的函数$F(x, y)$的极大点就是那个使$f(x, y)$，对曲线Γ的邻点来说，达到局部极大的点。但是我们可以证明——而这是在所有分析教本中看得到的——$f(x, y)$在曲线Γ上的任何局部极大点(x_0, y_0)都可依上述方法求出，只要在这一点上导数$\varphi'_x(x_0, y_0)$，$\varphi'_y(x_0, y_0)$不同时变为零[1]。

1) 在斯米尔诺夫《高等数学教程》(卷一第二分册第 427 页)中读者可以找到一个简单的例子，说明我们若机械地应用拉格朗日方法而除了上面所列举的点外不去考虑那些既满足(45)又满足条件$\varphi'_x(x_0, y_0)=0$，$\varphi'_y(x_0, y_0)=0$的点，则由于最后这个特征，就求不出解答来。

让我们把在这一小节开头所引述的例子用拉格朗日方法来求解. 在这一场合 $f(x, y)=xy$; $\varphi(x, y)=x^2+y^2-4R^2$. 构成方程 (43), (44), (45):

$$y+2\lambda x=0, \quad x+2\lambda y=0, \quad x^2+y^2=4R^2.$$

考虑到 x 与 y 都是正的, 由这方程组得到唯一的解:

$$x=y=R\sqrt{2} \quad \left(\lambda=-\frac{1}{2}\right).$$

对于这样彼此相等的 x 与 y, 也就是说, 在内接正方形的场合, 面积确实达到极大.

拉格朗日方法还可推广到三元以及更多元函数的场合. 类似条件(45)的辅助方程在这时可以有好几个(但少于元数), 并且就得引入相应数目的辅助乘数.

这里是关于寻求相对极大或极小问题的另外两个例子.

例 1 要制造一个有已知容积 V 的无盖圆桶, 在怎样的高度 h 及怎样的半径 r 时所用的铁片材料最省? 也就是说, 桶的表面积为最小?

这问题显然可归结于寻求 r 与 h 二元函数

$$f(r, h)=2\pi rh+\pi r^2$$

的极小, 但附有条件 $\pi r^2h=V$, 这条件可写成

$$\varphi(r, h)=\pi r^2h-V=0.$$

例 2 有一个动点须从 A 前进至 B(图 32). 在 AM 一段路程上它以速度 v_1 移动, 而在 MB 一段路程上以速度 v_2 移动. 问点 M 应当放在直线 DD' 的哪一点上, 这动点才能最迅速地自 A 至 B 通过全程?

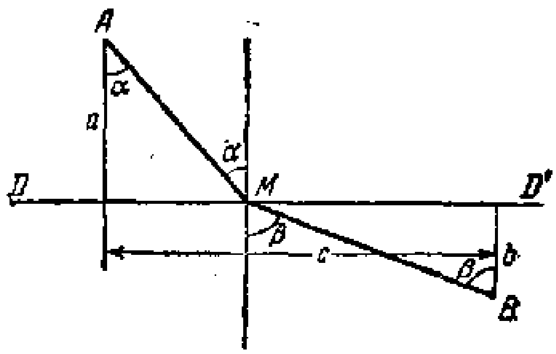


图 32

让我们取已在图 32 中标出的角 α 与 β 作为未知数. 自点 A 与 B 向直线 DD' 所作垂直线的长度 a 与 b 以及这两个垂足之间

的距离 c 都是我们所已知的．容易看出，动点通过全程的时间可由公式

$$f(\alpha, \beta)=\frac{a}{v_1 \cos \alpha}+\frac{b}{v_2 \cos \beta}$$

表出．我们要把这个表达式的极小找出来，但不能忘记在 α 与 β 之间有关系式

$$a \operatorname{tg} \alpha+b \operatorname{tg} \beta=c.$$

所引述的两个例子读者都可以自己应用拉格朗日方法来求解．在第二个例子中容易看到，当点 M 在最有利的位置时有条件

$$\frac{\sin \alpha}{\sin \beta}=\frac{v_1}{v_2}$$

成立．这就是著名的光线折射定律．这样，当光线从一种介质射至另一种介质时，它所受到的屈折总使它以最短的时间从一种介质的一点到达另一种介质的一点．这一类的结论已经不仅有计算上的，而且还有极大的认识上的趣味；它们促使精确的自然科学去进窥自然界的更深奥和普遍的规律性．

最后，我们要指出，在用拉格朗日方法解题时所引入的乘数 λ 并非仅仅是一些辅助的数．它们是每次跟各个特殊问题的本质有密切联系的，并且具有跟该问题有关联的具体意义．

§13．积分概念的推广

在§10中我们把和式

$$\sum_{i=1}^{n} f(\xi_i) \Delta x_i$$

当区间 $[a, b]$ 的最大分段 Δx_i 趋于零时的极限称为函数 $f(x)$ 在区间 $[a, b]$ 上的定积分．尽管确实有这种极限存在的函数 $f(x)$ 是包罗很广的一类(可积函数类)，特别是一切连续函数以及甚至于许多间断函数也都包括在内，但这类函数有着严重的缺点．把两个可积函数相加、相减、相乘以及在一定条件下相除，我们可以证明所得到的函数又是可积的．对于 $\frac{f(x)}{\varphi(x)}$，只要量 $\frac{1}{\varphi(x)}$ 在 $[a, b]$

中有界，这是无论如何正确的。但若我们从一连串可积函数 $f_1(x)$, $f_2(x)$, $f_3(x)$, … 经过取极限的步骤而得到某一函数，使得

$$f(x)=\lim_{n\to\infty} f_n(x),$$

对于 $[a, b]$ 中的一切 x 成立，则极限函数 $f(x)$ 未必一定是可积的。

在许多场合，上述这种情况以及其他一些情况，都在广泛应用极限过程的数学工具本身中造成异常的错综复杂。

这种处境的出路是在积分概念的进一步推广中找到的。在这样的推广中最重要的就是勒贝格积分，读者可以在讲述实变函数论的第三卷第十五章中看到。我们在那里要说一说实践上极重要的关于积分在其他方面的推广。

重积分 我们已经熟悉了在一维区域——区间——上所给定的一元函数的积分过程。但类似的过程还可能推广到在相应区域上所给定的二元、三元以及一般地说多元函数。

例如，设在直角坐标系中已给定曲面

$$z=f(x, y),$$

而在 Oxy 平面中已给定由闭曲线 Γ 所围成的区域 G。我们要确定由这个曲面、Oxy 平面及通过曲线 Γ 而母线平行 Oz 轴的柱面所围成的体积（图33）。为了要解决这个问题，我们用任何平行于 Ox 与 Oy 轴的直线网把平面区域 G 划成小的分区，并把其中成为完全矩形的分区逐一编号：

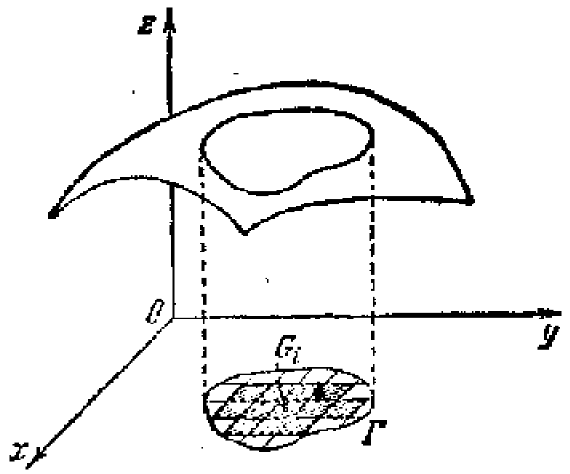


图 33

$$G_1, G_2, \cdots, G_n.$$

若直线网充分地细密，则区域 G 的绝大部分已被编了号的矩形所占尽，我们在每一个这样的矩形中选定任意一点：

$$(\xi_1, \eta_1), (\xi_2, \eta_2), \cdots, (\xi_n, \eta_n),$$

并且为简便起见，不仅用 G_i 记矩形而且还用 G_i 记这矩形面积的

数值. 于是让我们构成和式

$$S_n=f(\xi_1,\ \eta_1)G_1+f(\xi_2,\ \eta_2)G_2+\cdots+f(\xi_n,\ \eta_n)G_n$$

$$=\sum_{i=1}^{n}f(\xi_i,\ \eta_i)G_i. \tag{47}$$

显然,若曲面是连续的而直线网充分地细密,则我们可以使这个和式跟所求体积 V 任意地接近. 若在分区划得愈来愈小(即连各矩形对角线中的最大的也趋于零)时, 取和式(47)的极限, 我们就准确地求得了我们的体积:

$$\lim_{\max d(G_i)\to 0}\sum_{i=1}^{n}f(\xi_i,\ \eta_i)G_i=V. \tag{48}$$

由分析的观点看来,要确定体积 V, 就需要对函数 $f(x,\ y)$ 及它的定义域 G 作等式(48)左面所指出的某种数学运算. 这种运算称为 f 在区域 G 中的积分运算, 而运算的结果是函数 f 在区域 G 上的(二重)积分. 这结果通常是这样记写的:

$$\iint_G f(x,\ y)dx\,dy=\lim_{\max d(G_i)\to 0}\sum_{i=1}^{n}f(\xi_i,\ \eta_i)G_i. \tag{49}$$

类似地我们可以定义三元函数在三维区域 G 上的积分, 这个区域 G 是空间的某一种形体. 我们又把区域 G 划成小的分区,但这次所用的,是跟空间坐标平面相平行的平面,从这些分区中间挑出完全的长方体并逐一编号:

$$G_1,\ G_2,\ \cdots,\ G_n.$$

在每一个长方体中取一个任意点:

$$(\xi_1,\ \eta_1,\ \zeta_1),\ (\xi_2,\ \eta_2,\ \zeta_2),\ \cdots,\ (\xi_n,\ \eta_n,\ \zeta_n),$$

并构成和式

$$S=\sum_{i=1}^{n}f(\xi_i,\ \eta_i,\ \zeta_i)G_i, \tag{50}$$

其中 G_i 就表示长方体 G_i 的体积. 最后,我们把 $f(x,\ y,\ z)$ 在区域 G 上的积分定义为和式(50)当最大直径 $d(G_i)$ 趋于零时的极限:

$$\lim_{\max d(G_i)\to 0}\sum_{i=1}^{n}f(\xi_i,\ \eta_i,\ \zeta_i)G_i=\iiint_G f(x,\ y,\ z)dx\,dy\,dz. \tag{51}$$

让我们来看一个例子. 试想象区域 G 充满了一种不均匀的

质量，而且表达质量在 G 中分布密度的函数 $\rho(x, y, z)$ 是已知的. 质量在点 (x, y, z) 的密度 $\rho(x, y, z)$ 被定义为：包含点 (x, y, z) 在内的任何小区域的质量与其体积之比，当这区域的直径[1] 趋于零时的极限. 为了要确定物体 G 的质量，我们自然要推论如下. 用一连串跟坐标平面平行的平面把区域 G 划成小的分区，并把其中成为完全长方体的逐一编号：

$$G_1, G_2, \cdots, G_n.$$

若划分 G 的平面充分地细密，则我们即使把非完全长方体区域的质量略去而把每一个正则区域 G_i(完全长方体)的质量近似地由下列乘积来确定：

$$\rho(\xi_i, \eta_i, \zeta_i)G_i,$$

其中 (ξ_i, η_i, ζ_i) 是 G_i 的任意一点，也将招致很小的误差. 结果质量 M 的近似值将由和式

$$S_n=\sum_{i=1}^{n}\rho(\xi_i, \eta_i, \zeta_i)G_i$$

表出，而其准确的表达式显然是这一和式当 G_i 的最大直径趋于零时的极限，即

$$M=\iiint_G \rho(x, y, z)\,dx\,dy\,dz=\lim_{\max d(G_i)\to 0}\sum_{i=1}^{n}\rho(\xi_i, \eta_i, \zeta_i)G_i.$$

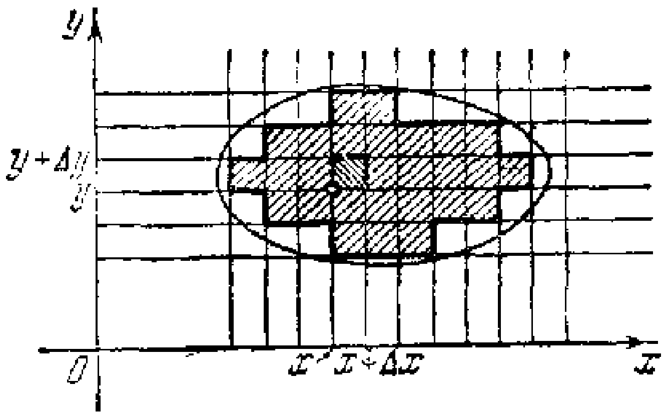


图 34

积分(49)与(51)分别有二重与三重积分之称.

让我们来考察一个导致二重积分的问题. 试想象在平面上有水流过. 同时，在平面的不同地点有地下的水以不同的强度 $f(x, y)$ 渗出来(或者，反之，有水流入土壤). 让我们选定一个由闭合周界所围

1) 一个区域的两点之间距离的上确界称为这区域的直径.

成的区域 G(图 34)，并且假定我们已知强度 $f(x, y)$，即对于区域 G 的每一点已知每分钟每平方公分平面上所渗出的地下水的数量(在地下水渗出的地方，$f(x, y)>0$，而在水从土壤中流走的地方，$f(x, y)<0$)．问在整个区域 G 中每分钟渗出多少水？

若我们把区域 G 划成不大的分区，在每一分区上把 $f(x, y)$ 当作固定不变来近似地计算渗出的水量，然后在分区划得愈来愈小时取极限，则我们就得到土壤所放出的总水量的积分式：

$$\iint_G f(x, y)\,dx\,dy.$$

最先引用二重积分的乃是欧拉．现在在各种极不相同的计算与研究中二重积分已成为经常使用的工具．

我们可以证明——但这不在我们的讨论范围之内——重积分的计算一般地可以化成寻常一维积分的叠次计算．

线积分与面积分 最后，我们应当指出，积分还可以有另外两种推广．例如，要确定对于沿着已知曲线运动的质点所施变力的功，自然地引导到所谓曲线积分；而要寻求以已知面密度连续地分布着电的曲面上的总电荷，则引导到一种新的概念——曲面积分．

设在空间有液体流过(图 35)而同时液体粒子在点 (x, y) 的速度由函数 $P(x, y)$ 表出．这样，速度是跟 z 无关的．若我们要想

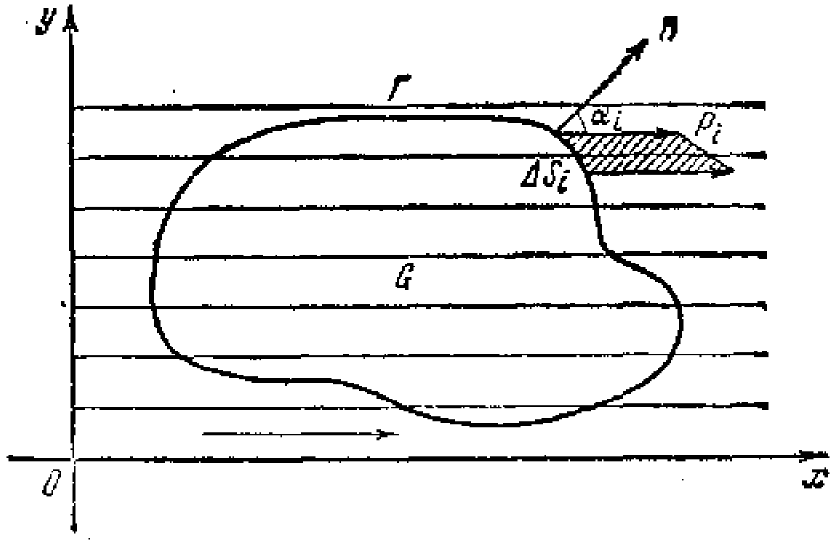


图 35

确定每分钟流出周界 Γ[1] 之外的液体数量，则可推论如下．把 Γ 划分为小段 Δs_i，流过一小段 Δs_i 的水量就近似地等于图 35 中加有阴影的，好像是在一分钟内从周界这一段挤出来的一小股液体．但加有阴影的平行四边形的面积等于

$$P_i(x,\ y)\cdot\Delta s_i\cdot\cos\alpha_i,$$

式中 α_i 是 x 轴的方向 $\bar{x}$ 与从(由周界 Γ 所围成的)区域 G 向外所作的切线方向(沿着这切线就可以近似地当作是曲线段 Δs_i 的方向)的垂线 $\bar{n}$ 之间的夹角．把这样的平行四边形面积一一相加，并且在周界 Γ 愈分愈细时取极限，我们就得到一分钟内流过周界 Γ 的水量，记作

$$\int_\Gamma P(x,\ y)\cos(\bar{n},\ \bar{x})ds,$$

称为曲线积分．若水流不是平行的，则流速将在每一点 $(x,\ y)$ 有沿 x 轴的分量 $P(x,\ y)$ 及沿 y 轴的分量 $Q(x,\ y)$．在这一场合，可以用类似的推论来确定流过周界 Γ 的水量将等于

$$\int_\Gamma [P(x,\ y)\cos(\bar{n},\ \bar{x})+Q(x,\ y)\cos(\bar{n},\ \bar{y})]ds^{2)}.$$

当我们谈到在曲面 G 上对已在这曲面各点 $M(x,\ y,\ z)$ 上给定的函数 $f(M)$ 求积分时，这就是指下列和式在积分域 G 被愈来愈细地划成面积等于 $\Delta\sigma_i$ 的分区时所趋极限：

$$\lim\sum_{i=1}^{n} f(M_i)\Delta\sigma_i=\iint_G f(x,\ y,\ z)d\sigma.$$

对于重积分、线积分与面积分，都不仅有普遍的计算与变换方法，而且也有近似算法．

奥斯特洛格拉特斯基公式 在某一体积(或平面区域)上的三重(或二重)积分与这体积表面(或区域边界)上的面积分(或线积

---

1) 更准确地说，流过底为周界 Γ 而高等于 1 的柱面．

2) 因为在曲线上作很小的位移时坐标 y 的微分恰巧等于 $\cos(\bar{n},\ \bar{x})ds$ 而微分 dx 等于 $\cos(\bar{n},\ \bar{y})ds$，所以最后这个积分常常写成下面的形式：

$$\int_\Gamma [P(x,\ y)dy-Q(x,\ y)dx].$$

分）之间有着极重要的相倚性，这种相倚性的极普遍形式是在上一世纪中叶被奥斯特洛格拉特斯基发现的.

我们不准备在这里证明这个具有广泛应用的奥斯特洛格拉特斯基一般公式，但力图在最简单的特殊场合中用例子来说明这个公式.

试想像在土壤的平面上有稳定的水流通过，其中始终有一部分水漏入土壤或从土壤里出来. 让我们选定某一个由周界 Γ 所围成的区域 G，并假定对于这区域的每一点已知水流速度在 x 轴与 y 轴的分量 $P(x, y)$ 与 $Q(x, y)$.

让我们计算在点 (x, y) 的附近水以怎样的强度从土壤里出来. 为此我们要考虑一个紧靠着点 (x, y) 而两边为 Δx 与 Δy 的小矩形.

由于速度分量 $P(x, y)$ 在一分钟内通过左纵界而流入这小矩形的水约为 $P(x, y)\Delta y$，而同时通过右纵界而流出的水约为 $P(x+\Delta x, y)\Delta y$. 整个说来，在单位面积上通过小正方形左右纵界而流出的水量近似地等于

$$\frac{[P(x+\Delta x, y)-P(x, y)]\Delta y}{\Delta x\,\Delta y}.$$

若令 Δx 趋于零，则在极限中得到

$$\frac{\partial P}{\partial x}.$$

相应地，水在 y 轴方向流出这小区域的强度由量

$$\frac{\partial Q}{\partial y}$$

表出. 这就是说，在点 (x, y) 上水从土壤渗出的强度将等于

$$\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}.$$

从土壤里出来的水的总量，像我们在以前所已经看到的那样，就等于这个表达每点上渗水强度的函数的二重积分，即

$$\iint\limits_G\left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}\right)dx\,dy. \tag{52}$$

所有这些水量应当同时流出周界 Γ 之外．但流出周界 Γ 的总水量，我们已经知道，由 Γ 上的线积分

$$\int_{\Gamma}[P(x, y)\cos(\bar{n}, \bar{x})+Q(x, y)\cos(\bar{n}, \bar{y})]ds \tag{53}$$

表出．

量(52)与(53)的相等也就表达了最简单的二维场合的奥斯特洛格拉特斯基公式

$$\iint_{G}\left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}\right)dx\,dy$$

$$=\int_{\Gamma}[P(x, y)\cos(\bar{n}, \bar{x})+Q(x, y)\cos(\bar{n}, \bar{y})]ds.$$

我们只用物理上的例子说明了这个公式的意义．它是可以在数学上证明的．

这样，奥斯特洛格拉特斯基的数学定理反映出现实的确定的规律性，而在上面的例子中我们就把它理解为不可压缩流体在数量上的保持平衡．

奥斯特洛格拉特斯基还证明了更普遍的、把多维体积上的积分跟这体积边界上的积分联系起来的公式．特别是对于由曲面 Γ 所围成的三维形体 G，这公式有下面的样子：

$$\iiint_{G}\left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}\right)dx\,dy\,dz$$

$$=\iint_{\Gamma}[P\cos(\bar{n}, \bar{x})+Q\cos(\bar{n}, \bar{y})+R\cos(\bar{n}, \bar{z})]d\sigma,$$

式中 $d\sigma$ 是面元素．

值得注意的是积分学基本定理

$$\int_{a}^{b}f(x)dx=F(b)-F(a) \tag{54}$$

可以看作是奥斯特洛格拉特斯基公式的一维场合．等式(54)把区间 $[a, b]$ 上的积分跟由两个端点 a 与 b 所构成的“零维”边界上的“积分”联系了起来．

公式(54)可以用下面的比拟来加以说明．试想象有水在一个

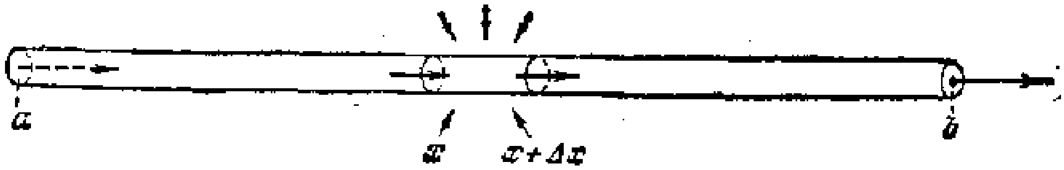


图 36

截面为常数 $s=1$ 的直管中流过，而水流的速度 $F(x)$ 随截面而异(图 36)．通过多孔的管壁在不同的截面有水以不同的强度从外面渗入管中(或从管中渗出)．若我们考虑直管从 x 到 $x+\Delta x$ 的一段，则在单位时间内从这一段渗入管中的水量应当跟沿管流出与流入这一段的水量之差 $F(x+\Delta x)-F(x)$ 相抵．所以在这一段渗入的水量跟差 $F(x+\Delta x)-F(x)$ 相同，而强度 $f(x)$，即从无限小的一段所渗入的水量与该段距离之比，就等于：

$$f(x)=\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}=F'(x).$$

在 $[a,b]$ 整个一段上流入管中与流出管外的总水量应当是相同的．但通过管壁而流入的量等于 $\int_a^b f(x)dx$，而沿管通过两头而流出的量是 $F(b)-F(a)$．公式(54)恰巧表达了这些量的相等．

§14．级　　数

级数概念　在数学中像下面这样的表达式：

$$u_0+u_1+u_2+\cdots$$

称为级数．

数 u_k 称为级数的项．级数的项无穷多并且都按照一定的次序排列，因此相应于每一个自然数 k 有一个确定的值 u_k．

读者应当注意，我们现在不讲能不能计算及怎样计算这种表达式．在表达式的各项 u_k 之间都有着加号，这一情况好像指出所有的项必须加起来．但是它们的数目无穷多，而数的加法只是就有尽个数的被加数来定义的．

让我们把级数最初 n 项的和记作 S_n；我们称它为级数的前 n 项的部分和．结果我们得到数列：

$$
\begin{aligned}
&S_1=u_0,\\
&S_2=u_0+u_1,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&S_n=u_0+u_1+\cdots+u_{n-1},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

并且可以说这里是一个变量 S_n, 其中 $n=1, 2, \cdots$.

若变量 S_n 在 $n\to\infty$ 时趋于一确定而有尽的极限:

$$\lim_{n\to\infty} S_n=S,$$

则级数称为收敛. 这一个极限便称为级数的和. 在这种场合我们写成

$$S=u_0+u_1+u_2+\cdots.$$

若在 $n\to\infty$ 时 S_n 的极限不存在, 则级数称为发散,而在这种场合再要说级数的和就没有意义了[1]. 不过, 若所有的 u_n 有相同的正负号,则在这种场合我们通常说级数的和等于正或负的无穷大.

作为一个例子,让我们来看一看公比为 x 的几何级数:

$$1+x+x^2+\cdots,$$

它的最初 n 项之和等于

$$S_n(x)=\frac{1-x^n}{1-x}\quad (x\neq 1); \tag{55}$$

在 $|x|<1$ 的场合,这个和有极限:

$$\lim_{n\to\infty} S_n(x)=\frac{1}{1-x},$$

因此当 $|x|<1$ 时可以写成

$$\frac{1}{1-x}=1+x+x^2+\cdots,$$

若 $|x|>1$, 则显然

$$\lim_{n\to\infty} S_n(x)=\infty,$$

---

1) 我们要指出,级数和还可以有推广的定义,根据这种定义我们可以或多或少自然地把"广义和"的概念归属于发散级数. 这样的级数就称为可求和的. 以发散级数的广义和来作运算,有时候是有效用的.

而级数就发散了. 当 $x=1$ 时情况也相同, 这可以直接证实, 因为公式(55)在 $x=1$ 时已经没有意义. 最后, 当 $x=-1$ 时级数的前 n 项的和轮流地取值 $+1$ 与 0, 所以级数也将发散.

相应于每一个级数, 其部分和的诸值 $S_1, S_2, S_3, \cdots$ 构成一个确定的序列. 级数的收敛与否就决定于这序列是否有极限. 反之, 相应于任何的数的序列 $S_1, S_2, S_3, \cdots$, 就有级数:

$$S_1+(S_2-S_1)+(S_3-S_2)+\cdots,$$

其部分和便是这序列中的各数. 这样, 按序列取值的变量的理论可以归结为相应的级数的理论, 而反之亦然. 但是这两种理论各自有其独立的意义. 在有些场合以直接讨论变量比较便利, 而在另外一些场合则以考虑跟它等价的级数比较便利.

我们要指出, 级数从很早的时候起就已经成为一种重要的工具, 借以计算和表达种种不同的量, 首先是函数. 自然, 数学家对于级数概念的见解在历史上是有变更的, 并且跟整个无穷小分析的发展情形有紧密联系. 上述关于级数收敛与发散的清晰定义是在上世纪初才同跟它有关的极限概念一起形成的.

若级数收敛, 则其普通项在 n 无限制增大时趋于零, 因为

$$\lim_{n\to\infty} u_n=\lim_{n\to\infty}(S_{n+1}-S_n)=S-S=0.$$

从后面的一些例子, 我们将看到, 反过来的论断一般地说是不正确的. 但所说准则很有用, 因为它提供了级数收敛的**必要**条件. 例如, 公比为 $x>1$ 的几何级数是发散的, 这一情况从级数的普通项不趋于零已可断定.

若级数由正项构成, 则它的部分和 S_n 随着 n 而增大, 并且只可能有两种场合: 或者变量 S_n 在 n 充分大时成为大于任何预先给定的数 A, 这时对于 n 的随后诸值 S_n 就变成永远大于 A, 因而 $\lim_{n\to\infty} S_n=\infty$, 就是说, 级数发散; 或者有这样一个数 A 存在, 对于 n 的一切值 S_n 不超过 A, 但这时变量 S_n 将不得不趋于一个确定、有限而不超过 A 的极限, 因而我们的级数将是收敛的.

级数的收敛问题 关于一个已给的级数是否收敛或发散的问

题，往往可以借助于跟另外一个级数的比较而得到解决．在这方面，我们所通常利用的是下述准则．

若已知两个正项级数

$$u_0+u_1+u_2+\cdots,$$

$$v_0+v_1+v_2+\cdots,$$

并且对于某一个值以后的一切自然数 n，有不等式

$$u_n\leqslant v_n$$

成立，则当第二个级数收敛时第一个也收敛，而当第一个级数发散时第二个也发散．

作为一个例子，让我们来看级数

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+\cdots,$$

这就是所谓调和级数．它的各项相应地不小于级数

$$1+\underline{\frac{1}{2}}+\underline{\frac{1}{4}+\frac{1}{4}}+\underline{\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}}+\underbrace{\frac{1}{16}+\cdots+\frac{1}{16}}_{8\text{项}}+\cdots$$

的各项，式中在一条横线上的各项之和都等于 $\frac{1}{2}$．

显然，第二个级数的部分和 S_n 随着 n 而无限制增大，因此调和级数是发散的．

级数

$$1+\frac{1}{2^\alpha}+\frac{1}{3^\alpha}+\frac{1}{4^\alpha}+\cdots, \qquad (56)$$

式中 α 是小于 1 的正数，显然也发散，因为对于任何 n 有

$$\frac{1}{n^\alpha}>\frac{1}{n} \quad (0<\alpha<1).$$

另一方面，可以证明级数(56)在 $\alpha>1$ 时就将是收敛的．我们在这里只就 $\alpha\geqslant 2$ 的情形来证明；为此让我们来考虑正项级数

$$\left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\cdots+\left(\frac{1}{n-1}-\frac{1}{n}\right)+\cdots,$$

这级数是收敛的，而且其和等于 1，因为其部分和

$$S_n=1-\frac{1}{n+1}\to 1 \quad (n\to\infty).$$

另一方面，其普通项满足不等式：

$$\frac{1}{n-1}-\frac{1}{n}=\frac{1}{(n-1)n}>\frac{1}{n^2},$$

由此可知 $$1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots$$

是收敛的．在 $\alpha>2$ 时级数(56)越发是收敛的了．

我们还要引述一个常常使用的关于审定正项级数收敛与发散的准则，称为达朗贝尔准则，但不加以证明．

假定比式 $\frac{u_{n+1}}{u_n}$ 在 n 趋于无穷大时有极限 q，于是在 $q<1$ 时级数显然是收敛的，在 $q>1$ 时显然是发散的，但在 $q=1$ 时，级数收敛与否的问题依然悬而未决．

由有限个被加数所组成的通常和式并不会因为各数地位的交换而有所改变．但对于无穷级数，一般地说，这是不正确的．确有一些收敛级数，只要把其中各项的位置交换，就可能使它的和发生变化或甚至于使它变为发散级数．一个级数，若具有这样不稳固的和，已失掉了通常和式的一种主要性质——可交换性．所以把保有这种性质的级数区别出来是极其重要的．这就是所谓绝对收敛级数．我们说级数

$$u_0+u_1+u_2+u_3+\cdots$$

绝对收敛，意思就是说，由其各项的绝对值所组成的级数

$$|u_0|+|u_1|+|u_2|+|u_3|+\cdots$$

是收敛的．可以证明，绝对收敛级数总是收敛的，也就是说，它的部分和 S_n 趋于有限的极限．任何收敛的级数，要是各项有着相同的正负号，显然就是绝对收敛的．

级数 $$\frac{\sin x}{1^2}+\frac{\sin 2x}{2^2}+\frac{\sin 3x}{3^2}+\cdots$$

可以作为绝对收敛级数的一个例子，因为级数

$$\left|\frac{\sin x}{1^2}\right|+\left|\frac{\sin 2x}{2^2}\right|+\left|\frac{\sin 3x}{3^2}\right|+\cdots$$

的各项不超过收敛级数

$$1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots$$

的相应各项.

可以作为收敛而非绝对收敛级数的例子的是下面的级数:

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots,$$

读者可试自己加以证明.

函数项级数. 均匀收敛级数 在分析中常常要碰到各项都由 x 的函数所构成的级数. 在上一小节中我们就看到过这种级数的例子,如级数 $1+x+x^2+x^3+\cdots$. 它对于 x 的某些值是收敛的,对于另外一些值是发散的. 那些对于某一区间(这区间也可以是整个实数轴或半轴)内一切 x 值都收敛的函数项级数形成了一种在应用上很重要的场合. 我们有必要对这样的级数逐项求导数,逐项求积分,并阐明其和是否连续等等问题. 若问题在于由有限个项所组成的通常和式,则我们已有简单普遍的法则. 我们知道可微函数之和的导数等于各函数的导数之和,连续函数之和的积分就是各函数的积分之和,连续函数之和也是连续函数——这一切都是对有限个项说的.

但在转到无穷级数时这些简单的法则,一般地说,不再有效了. 我们可以举出收敛函数项级数的许多例子,对它们逐项求积分与求导数的法则都是不正确的. 恰巧同样地,当级数由连续函数项组成时,其和可以不是连续函数. 另一方面,也有许多级数对于这些法则的效应正像通常有限项的和一样.

对于这一问题的相应的深刻研究显示了这些法则的应用是可以预先保证合法的,只要所考虑的级数不仅在所考虑的区间(x 的变化区间)的每一个个别点上收敛,而且这种收敛是对整个区间均匀的. 这样,在数学分析中(于十九世纪中叶)就形成了级数均匀收敛的重要概念.

让我们来考虑级数

$$S(x)=u_0(x)+u_1(x)+u_2(x)+\cdots,$$

其各项都是定义于区间 $[a, b]$ 的函数. 我们还要假定,对于这区

间的每一个个别 x 值，它收敛于某一个和 $S(x)$. 这级数的最初 n 项的和

$$S_n(x)=u_0(x)+u_1(x)+\cdots+u_{n-1}(x)$$

也是 x 的定义于$[a, b]$的某一个函数.

现在让我们来考察量 η_n，它等于 $|S(x)-S_n(x)|$ 的值当 x 在 $[a, b]$上变动时的上确界[1]. 这个量我们这样记写:

$$\eta_n=\sup_{a\leq x\leq b}|S(x)-S_n(x)|^{2)}.$$

在量 $|S_n(x)-S(x)|$ 取得其极大值的场合(这种场合显然是有的，比如说，当 $S(x)$与 $S_n(x)$都是连续的时候)，η_n 就不过是 $|S(x)-S_n(x)|$在$[a, b]$上的极大值.

由于我们的级数对于区间 $[a, b]$ 的每一个个别 x 值都假定是收敛的，

$$\lim_{n\to\infty}|S(x)-S_n(x)|=0.$$

但这时量 η_n 可以趋于零. 也可以不趋于零. 若量 η_n 在 $n\to\infty$ 时趋于零，则级数称为**均匀收敛**. 在相反的场合，级数称为非均匀收敛. 我们可以在同样的意义上说到函数序列 $S_n(x)$ 的均匀与非均匀收敛，在这里不一定要把 S_n 当作部分和而跟级数联系起来.

例 1 函数项级数

$$\frac{1}{x+1}-\frac{1}{(x+1)(x+2)}-\frac{1}{(x+2)(x+3)}-\cdots,$$

我们假定只对于非负的 x 值，也就是说，在半直线$[0, \infty)$上有定义. 这级数可以写成下面的形式:

$$\frac{1}{x+1}+\left(\frac{1}{x+2}-\frac{1}{x+1}\right)+\left(\frac{1}{x+3}-\frac{1}{x+2}\right)+\cdots,$$

由此可知它的部分和等于

$$S_n(x)=\frac{1}{x+n},$$

而

$$\lim_{n\to\infty}S_n(x)=0,$$

1) 见第三卷第十五章.

2) sup 是拉丁文 superior(最高)的缩写.

所以这级数对于一切非负的 x 值收敛并有和 $S(x)=0$. 其次,

$$\eta_n=\sup_{0\leqslant x<\infty}|S(x)-S_n(x)|=\sup_{0\leqslant x<\infty}\frac{1}{x+n}=\frac{1}{n}\to 0 \quad (n\to\infty),$$

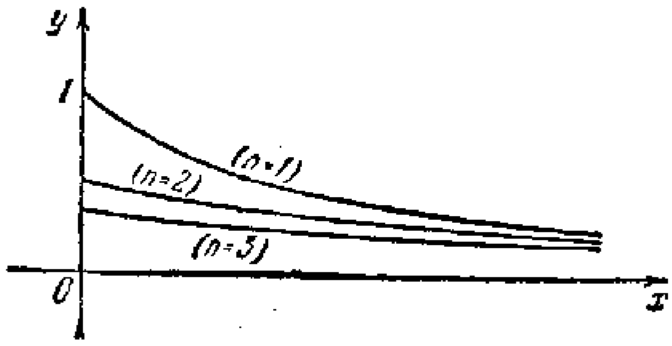


图 37

因而我们的级数在半轴 $[0, \infty]$ 上均匀收敛于零. 在图 37 中画出了部分和 $S_n(x)$ 的图形.

例 2 级数

$$x+x(x-1)+x^2(x-1)+\cdots$$

可以改写为

$$x+(x^2-x)+(x^3-x^2)+\cdots,$$

从而有

$$S_n(x)=x^n,$$

因此

$$\lim_{n\to\infty}S_n(x)=\begin{cases}0, & 当\ 0\leqslant x<1;\\ 1, & 当\ x=1.\end{cases}$$

这样, 级数的和是一个在区间 $[0, 1]$ 上间断的函数, 间断的地方在点 $x=1$. 量 $|S(x)-S_n(x)|$ 在 $[0, 1]$ 的每一个 x 上小于 1, 但在靠近 $x=1$ 的 x 点上, 就变得跟 1 任意接近. 所以对于一切 $n=1, 2, \cdots$ 有

$$\eta_n=\sup_{0\leqslant x\leqslant 1}|S(x)-S_n(x)|=1.$$

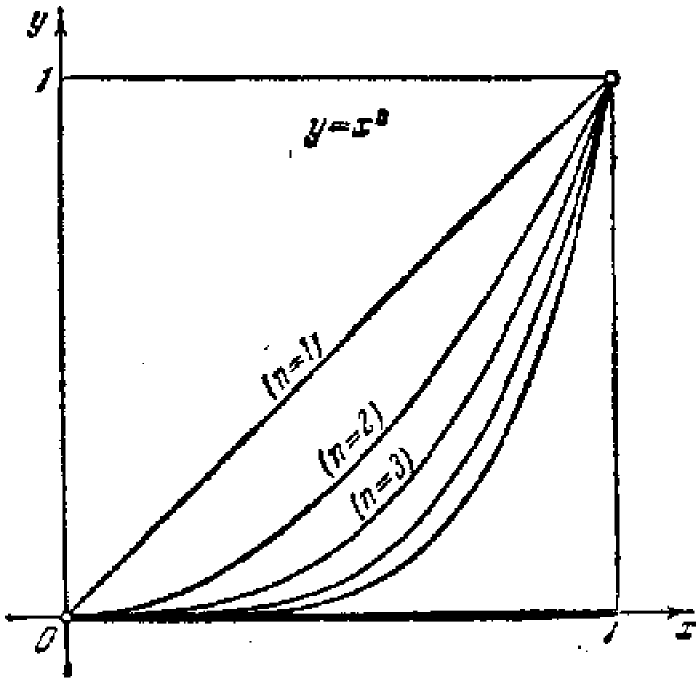


图 38

这样, 我们的级数在区间 $[0, 1]$ 上是非均匀收敛的. 在图 38 中绘出了函数 $S_n(x)$ 的图形. 这级数和的图形是由 x 轴上没有右端点的线段 $0\leqslant x<1$ 及点

(1, 1)构成的.

这个例子显示，由连续函数项组成的非均匀收敛级数之和实际上可以是一个间断函数.

另一方面，若就区间 $0\leqslant x\leqslant q$（其中 $q<1$）来考虑我们的级数，则

$$\eta_n=\sup_{0\leqslant x\leqslant q}|S(x)-S_n(x)|=\max_{0\leqslant x\leqslant q}x^n=q^n\underset{n\to\infty}{\longrightarrow}0,$$

因而在这区间上级数均匀收敛，而且我们看到级数的和也连续了. 这一情况，即均匀收敛级数之和为连续函数，像上文已经说过的那样，是一条可以严格地证明的普遍法则.

例 3 某一个级数的最初 n 项之和 $S_n(x)$ 的图形如图 39 中所绘出的粗折线. 显然，对一切 n 将有 $S_n(0)=0$; 若 $0<x\leqslant 1$，则对不小于 $\frac{1}{x}$ 的一切 n，即 $n\geqslant\frac{1}{x}$，将有 $S_n(x)=0$，因而对区间 [0, 1] 的任何 x 有

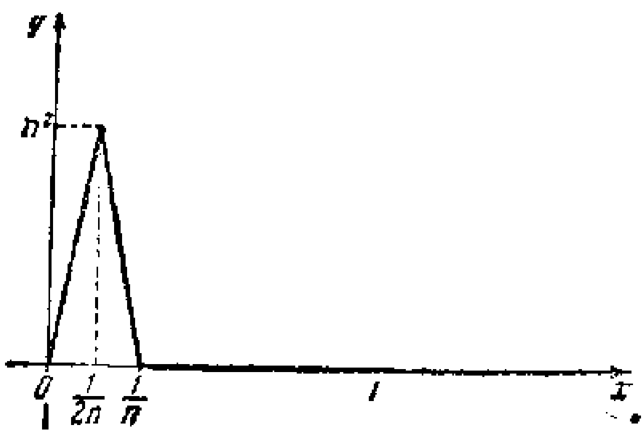


图 39

$$S(x)=\lim_{n\to\infty}S_n(x)=0.$$

另一方面，

$$\eta_n=\sup_{0\leqslant x\leqslant 1}|S(x)-S_n(x)|=\sup|S_n(x)|=n^2.$$

我们看到，量 η_n 不趋于零; 它甚至趋于无穷大. 我们要指出，跟所考虑的序列 $S_n(x)$ 相对应的级数不可能在区间 [0, 1] 上积分，因为

$$\int_0^1 S_1(x)\,dx=0,\quad \int_0^1 S_n(x)\,dx=\frac{1}{2}n^2\frac{1}{n}=\frac{n}{2},$$

而级数

$$\int_0^1 S_1(x)\,dx+\int_0^1[S_2(x)-S_1(x)]\,dx+\int_0^1[S_3(x)-S_2(x)]\,dx+\cdots$$

就变成这样一个发散级数了：

$$\frac{1}{2}+\left(\frac{2}{2}-\frac{1}{2}\right)+\left(\frac{3}{2}-\frac{2}{2}\right)+\left(\frac{4}{2}-\frac{3}{2}\right)+\cdots.$$

让我们把均匀收敛级数的主要性质不加证明而叙述如下：

1．在区间$[a, b]$上均匀收敛的连续函数项级数之和在这区间上连续．

2．若连续函数项级数

$$S(x)=u_0(x)+u_1(x)+u_2(x)+\cdots \tag{57}$$

在区间$[a, b]$上均匀收敛，则我们可以在这区间上把它逐项积分，就是说，对于$[a, b]$中的任何x_1, x_2有等式

$$\int_{x_1}^{x_2} S(t)dt=\int_{x_1}^{x_2} u_0(t)dt+\int_{x_1}^{x_2} u_1(t)dt+\cdots$$

成立．

3．设级数(57)在区间$[a, b]$上收敛而函数$u_k(x)$有连续导数，于是由级数(57)逐项求导数所得到的等式

$$S'(x)=u_0'(x)+u_1'(x)+u_2'(x)+\cdots \tag{58}$$

显然将在区间$[a, b]$上成立，只要在等式(58)中右面所得到的级数是均匀收敛的．

幂级数 在§9中我们说过，若定义于区间$[a, b]$的函数$f(x)$有任意阶导数，而且在$[a, b]$中任意一点x_0的充分小的邻域中可展开为向它自己收敛的泰勒级数：

$$f(x)=f(x_0)+\frac{f'(x_0)}{1}(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots, \tag{59}$$

则这函数称为分析函数．若引用记号

$$a_n=\frac{f^{(n)}(x_0)}{n!},$$

则这级数还可以写成下面的样子：

$$f(x)=a_0+a_1(x-x_0)+a_2(x-x_0)^2+\cdots. \tag{60}$$

任何这样的级数(其中a_1, a_2, $\cdots$是与x无关的常数)在数学中称为幂级数．

作为一个例子，让我们考虑幂级数

$$1+x+x^2+x^3+\cdots, \tag{61}$$

其各项构成一个几何级数.

我们知道, 对于区间 $-1<x<1$ 的一切 x 值这级数是收敛的, 而它的和等于

$$S(x)=\frac{1}{1-x}.$$

对于其余的 x 值该级数是发散的.

我们也容易看出, 这级数和与前 n 项的部分和之间的差由公式

$$S(x)-S_n(x)=\frac{x^n}{1-x} \tag{62}$$

表出, 而若 $-q\leqslant x\leqslant q$, 其中 q 是一个小于 1 的正数, 则

$$\eta_n=\max|S(x)-S_n(x)|=\frac{q^n}{1-q}.$$

由此可见, 在 n 无限制增大时, η_n 趋于零, 而我们的级数就在区间 $-q\leqslant x\leqslant q$ 上均匀收敛, 不论 q 是怎样一个小于 1 的正数.

容易验证, 函数

$$S(x)=\frac{1}{1-x}$$

的 n 阶导数等于

$$S^{(n)}(x)=\frac{n!}{(1-x)^{n+1}},$$

从而有

$$S^{(n)}(0)=n!,$$

由此可知, 函数 $S(x)$ 在 $x_0=0$ 的泰勒公式的前 n 项之和就跟级数 (59) 的前 n 项之和完全相同. 我们又知道, 公式的余项, 如等式 (62) 所表出, 对于区间 $-1<x<1$ 内的一切 x 在 n 无限制增大时趋于零. 于是证明了级数 (61) 就是其和 $S(x)$ 的泰勒公式.

我们还要指出一个事实. 在我们这级数的收敛区间 $-1<x<1$ 中取任意点 x_0, 容易看出, 对于跟 x_0 足够邻近的一切 x, 也就是说, 对于那些满足不等式

$$\frac{|x-x_0|}{1-x_0}<1$$

的一切 x, 下列等式总是正确的:

$$S(x)=\frac{1}{1-x}=\frac{1}{1-x_0}\frac{1}{\left(1-\frac{x-x_0}{1-x_0}\right)}$$

$$=\frac{1}{1-x_0}\left[1+\frac{x-x_0}{1-x_0}+\left(\frac{x-x_0}{1-x_0}\right)^2+\cdots\right]$$

$$=\frac{1}{1-x_0}+\frac{x-x_0}{(1-x_0)^2}+\frac{(x-x_0)^2}{(1-x_0)^3}+\cdots, \qquad (63)$$

读者无须特别费力就可以验证

$$\frac{S^{(n)}(x_0)}{n!}=\frac{1}{(1-x_0)^{n+1}}.$$

因此，级数(63)是其和 $S(x)$的泰勒级数，它在级数(61)收敛区间内任意点 x_0 的充分小邻域中收敛于 $S(x)$，这就是说，由于点 x_0 的任意性，$S(x)$是一个在这区间上的分析函数.

所有这些就特殊幂级数(61)而说的事实对于任意幂级数也都成立[1]. 这就是说，不论怎样一个形如(60)的幂级数，其中 a_k 是任意的、按任何规律分布的数，相应地有某一个非负数 R(特别是可以变为 ∞)，称为级数(60)的**收敛半径**，并且具有下述性质：

1. 对于所谓收敛区间 $x_0-R<x<x_0+R$ 内的一切 x 值，级数是收敛的，而它的和 $S(x)$在这区间上是 x 的分析函数. 这时在完全属于收敛区间的任何区间$[a, b]$上，收敛是均匀的. 这级数本身就是其和的泰勒级数.

2. 在收敛区间的两端，级数也许收敛，也许发散，得看级数的个别性质而定. 但在闭区间 $x_0-R\leqslant x\leqslant x_0+R$ 之外，它显然是发散的.

我们让读者去考察下列幂级数：

$$1+\frac{x}{1}+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots,$$

$$1+x+2!x^2+3!x^3+\cdots,$$

$$1+x+\frac{x^2}{2}+\frac{x^3}{3}+\cdots,$$

并验证第一个级数的收敛半径等于无穷大，第二个等于零，而第三

1) 在第二卷第九章中有较详细的叙述.

个等于一.

按照前面所给出的定义，每一个分析函数可以在其定义区间内任何点的充分小邻域中展开为收敛于这函数的幂级数. 反之，从以上所说的，可知每一个幂级数，只要其收敛半径不等于零，就在其收敛区间中有和是分析函数.

这样，我们看到幂级数是有机地跟分析函数相联系的. 我们还可以说，幂级数在其收敛区间上是表达分析函数的天然工具，而同时也就是以代数多项式来逼近分析函数的天然工具[1].

例如，由于函数 $\frac{1}{1-x}$ 可展开为在区间 $-1<x<1$ 内收敛的幂级数：

$$\frac{1}{1-x}=1+x+x^2+x^3+\cdots,$$

可知这级数在任何区间 $-a\leqslant x\leqslant a$ 上均匀收敛，只要 $a<1$，因此就有可能在整个区间 $[-a, a]$ 上借助于所考虑的级数的部分和来逼近这函数使其达到任何预期的准确度.

假定我们要在区间 $\left[-\frac{1}{2}, \frac{1}{2}\right]$ 上用多项式来逼近函数 $\frac{1}{1-x}$，使其准确到 0.01. 我们指出，对于这区间的一切 x 有下列不等式成立：

$$\left|\frac{1}{1-x}-1-x-\cdots-x^n\right|=|x^{n+1}+x^{n+2}+\cdots|$$

$$\leqslant|x|^{n+1}+|x|^{n+2}+\cdots\leqslant\frac{1}{2^{n+1}}+\frac{1}{2^{n+2}}+\cdots=\frac{1}{2^n},$$

而由于 $2^6=64$, $2^7=128$，可知所求的多项式，即在整个区间 $\left[-\frac{1}{2}, \frac{1}{2}\right]$ 上以 0.01 的准确度逼近所考虑函数的多项式，将有如下形式：

$$\frac{1}{1-x}\approx 1+x+x^2+\cdots+x^7.$$

我们还要提出幂级数的一个极有用的性质：在其收敛区间中它总是可以逐项求导数的. 这一性质在数学中被广泛地应用于求

1) 在幂级数收敛区间之外的近似表达，我们应用其他方法(见第二卷第十二章).

解各种问题.

例如，我们要在 $y(0)=1$ 的附加条件下求解微分方程 $y'=y$，可试求它的幂级数形式的解：

$$y=a_0+a_1x+a_2x^2+\cdots,$$

由于附加条件，必须令 $a_0=1$．假设这级数收敛，我们就有权对它逐项求导数；结果得到

$$y'=a_1+2a_2x+3a_3x^2+\cdots,$$

把这两个级数都代入微分方程并令等式两边的 x 的同幂项的系数相等，即得

$$a_k=\frac{1}{k!}\quad (k=1,\ 2,\ \cdots),$$

因而所求的解将有如下形式：

$$y=1+\frac{x}{1}+\frac{x^2}{2!}-\frac{x^3}{3!}+\cdots,$$

这级数已知是对一切 x 值收敛的，而其和等于 e^x，因而 $y=e^x$.

在这一场合级数的和是我们已知的初等函数．但并非经常如此：由解题所得到的收敛幂级数也可以有并非初等函数的和．由求解应用上很重要的贝塞尔微分方程所得到的级数

$$y_p(x)=x^p\left[1-\frac{x^2}{2(2p+2)}+\frac{x^4}{2\cdot 4(2p+2)(2p+4)}-\cdots\right]$$

便是一个例子．这样，幂级数又是形成新函数的工具了.

文　献

通俗小册子

В. Г. 保尔强斯基，什么是微分法?，商务印书馆，1957 年版.

А. И. Маркушевич, Ряды. Гостехиздат, 1947.

А. И. Маркушевич, Площади и логарифмы. Гостехиздат, 1952.

И. П. 那汤松，无穷小量的求和，科学出版社，1954 年版.

И. П. Натансон, Простейшие задачи на максимум и минимум. Гостехиздат, 1950.

系统的教程

A. Ф. 别尔曼特，数学解析教程，高等教育出版社，1955 年版。

B. И. 斯米尔诺夫，高等数学教程，卷一，高等教育出版社，1952 年版。

Г. П. Толстов, Курс математического анализа, т. I. Гостехиздат, 1954.

Г. М. 菲赫金哥尔茨，微积分学教程，高等教育出版社，1953—1955 年版。

A. Я. 辛钦，数学分析简明教程，高等教育出版社，1954 年版。

初等数学全书，第三卷，函数和极限，高等教育出版社，1955 年版。

(开头两篇包含有初等函数的研究，极限论的详细叙述，以及微积分学与级数论的初步知识。)

赵孟养 译

秦元勋 校

第三章 解析几何

§1. 绪　　论

在十七世纪前半叶，产生了数学的全新的一支，叫做解析几何，它使平面上的曲线与有两个未知数的代数方程之间建立了联系.

在数学中发生了极为稀有的情景：在一、二十年内出现了巨大的、全新的数学分支，并且这数学分支还是以非常简单的、但是一直未受到应有注意的观念为基础的．解析几何在十七世纪前半叶出现，决不是偶然的．当欧洲过渡到新的资本主义的生产方式时，有一系列的科学部门需要整个地加以改进．正当伽利略的学说和其他的一些学说开始奠定现代的力学时，在自然科学的所有领域里都累积了实验的数据，改善了观察方法，创立了新的理论来代替古老的玄学式的理论．在天文学方面，在先进的学说中间最后得胜的是哥白尼的学说．远洋航行的急骤发展急切地要求着天文学的知识和力学的原理.

力学在军事行动中也是必要的．作为圆锥截线的椭圆和抛物线，它们的几何性质早在离开当时将近 2000 年前的古希腊时代已经知道得很详细了；然而它们一直还像在希腊时代那样，只被当作几何学的对象．一到刻普勒发现行星沿椭圆轨道绕着太阳运动，伽利略发现抛出去的石子沿着抛物线的轨道飞去时，就必须要来计算这些椭圆，要来求出炮弹飞驰时所画出的抛物线了；就必须要来发掘由巴斯卡发现的大气压力随高度而递减的法则了；就必须要来实地算出各种不同物体的体积了，诸如此类不胜枚举.

所有这些问题在人类生活中几乎同时地引起了三门全新的数学科学的发展，它们就是：解析几何、微分法和积分法（包括解简单

的微分方程).

这三门新的学科从本质上改变了整个数学的面貌. 它们的运用使得直到那时还无法解决的问题变得容易解决了.

在十七世纪的前半叶, 一系列最优秀的数学家已经接近了解析几何的观念, 但是只有两位数学家才特别清楚地认识到创立新的数学部门的可能性. 其中一位是庞埃尔·弗尔马, 他是法国土鲁兹城的市议会的顾问, 世界上最卓越的数学家之一, 另一位是法国著名的哲学家蕊耐·笛卡儿. 解析几何的主要创立者无论如何总该推笛卡儿. 唯有作为哲学家的笛卡儿, 才提出了它的全面推广的问题. 笛卡儿发表了长篇的哲学论著"关于下列方法的讨论: 为了很好地在科学中指出其精神和发掘真理, 并应用于折射光学、气象学和几何学".

这著作的后一部分, 以"几何学"命名, 发表在 1637 年, 包含着我们现在叫做解析几何的那门数学理论的十分完全的(虽然有些杂乱的)叙述.

§2. 笛卡儿的两个基本观念

笛卡儿想要创造一种方法, 以便用来解决所有的几何问题, 给出这些问题的所谓一般的解法. 笛卡儿的理论以下面两个观念为基础: 坐标观念和利用坐标方法把带两个未知数的任意代数方程看成平面上的一条曲线的观念.

坐标观念 所谓平面上的点的坐标, 笛卡儿是指这个点的横坐标和纵坐标, 即这点到这平面上两条取定的互相垂直的直线(坐标轴)的距离(具有对应符号)的数值 x 和 y (参看第二章). 坐标轴的交点, 即有坐标(0, 0)的点, 叫做坐标原点.

笛卡儿坐标的引入促成了平面的所谓"算术化". 只要给定一对数 x, y, 就等于在几何上指出一个点, 而且反过来也对(图 1).

将带两个未知数的方程和平面上的曲线相对比的观念 笛卡儿的第二个观念如下. 在笛卡儿之前, 每当遇到带两个未知数的

代数方程 $F(x, y)=0$ 时，大家都说问题是不定的，从这个方程无法决定这两个未知数．这是因为其中的一个，例如 x 可以取任意的数，用这个数代替 x，就得到带一个未知数 y 的方程，一般说来从这个方程可以解出 y．然后这个任意取的 x 连同这样得到的 y 就满足给定的方程．所以对这种“不定的”方程，人们并不认为是值得关心的．

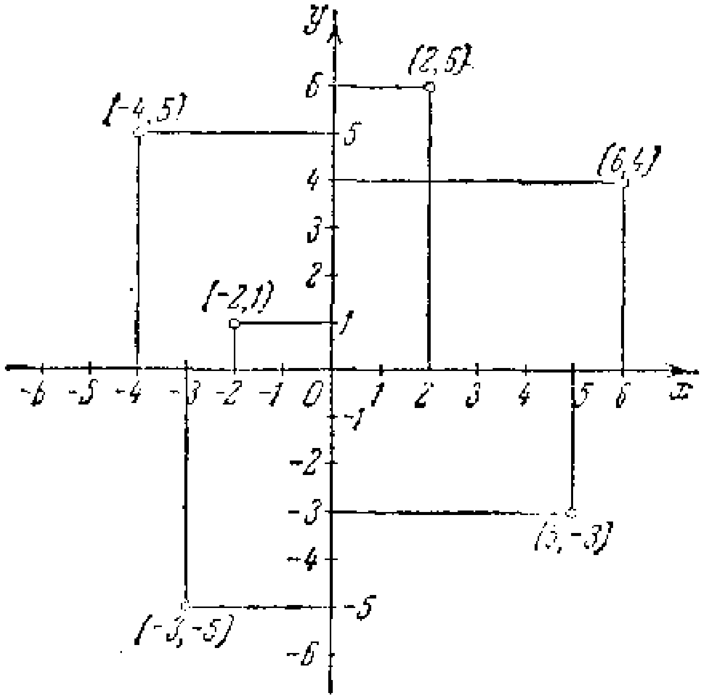


图 1

笛卡儿察觉到另一种情况，他设想带两个未知数的方程中的 x 是点的横坐标，与 x 对应的 y 是点的纵坐标．于是，如果让 x 连续地改变，则对于每个 x 都可以从方程算出完全确定的 y，因此一般地说也就得到了组成一条曲线的点集合(图 2)[1]．

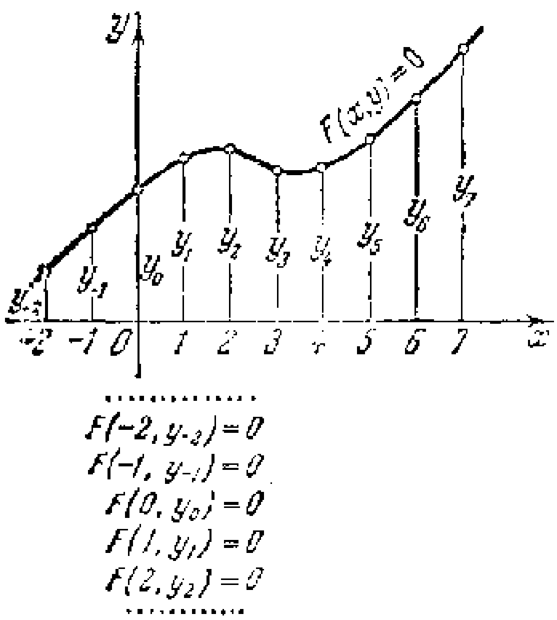


图 2

这样一来，对于带两个变数的每一个代数方程 $F(x, y)=0$，与它对应的是平面上的一条完全确定的曲线，那就是代表了平面上所有那些其坐标

1) 有时方程不被任何具实坐标的点 (x, y) 所满足，有时一共只有一个或者几个这样的点．在这种情况下，就说曲线是零曲线或者说曲线退化成点(参看第 221 页)。

满足方程 $F(x, y)=0$ 的点的集合的曲线.

笛卡儿的这种解释开创了整整一门新的科学.

解析几何所能解决的主要问题和解析几何的定义 解析几何造成了以下种种可能:

1) 通过计算来解决作图问题(例如分线段成已知比值, 参看本页);

2) 求由某种几何性质给定的曲线的方程(例如从到两个定点的距离之和是常数这个条件得到椭圆的方程, 参看第 209 页);

3) 用代数方法证明新的几何定理(例如推导牛顿的直径理论, 参看第 206 页);

4) 相反地, 从几何方面来看代数方程, 说明它的代数性质(例如, 利用抛物线与圆的交点来解三次和四次方程, 参看第 203—205 页).

因此, 解析几何是这样一个数学部门, 它在采用了坐标方法的同时, 运用代数方法来研究几何对象.

§3. 一些最简单的问题

分线段成已知比值的点的坐标 已知两个点 M_1 和 M_2 的坐标 (x_1, y_1) 和 (x_2, y_2), 求分线段 M_1M_2 成比值 m 比 n 的点 M 的坐标 (x, y) (图 3). 从加线条的三角形的相似性我们得到

$$\frac{x-x_1}{x_2-x}=\frac{m}{n},$$

于是

$$x=\frac{nx_1+mx_2}{m+n},$$

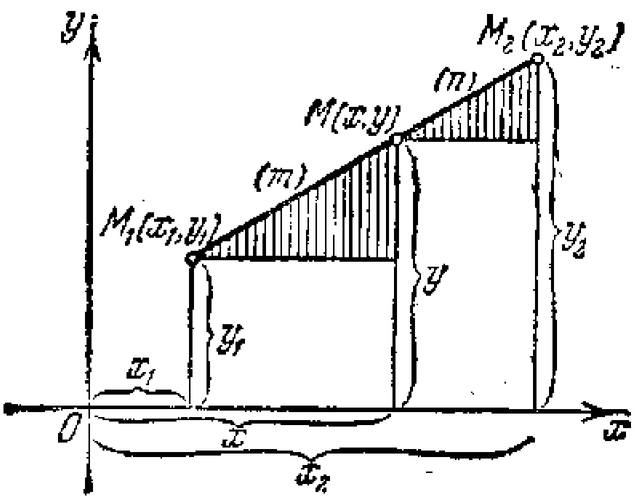


图 3

$$\frac{y-y_1}{y_2-y}=\frac{m}{n}, \quad 于是 \quad y=\frac{ny_1+my_2}{m+n}.$$

两点之间的距离 求坐标是(x_1, y_1)和(x_2, y_2)的两个点M_1和M_2之间的距离d. 根据毕达哥拉斯定理，从加线条的直角三角形(图 4) 得到

$$d=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}.$$

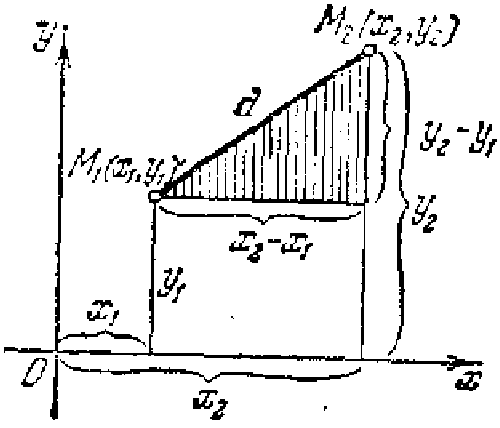


图 4

三角形的面积 求$\triangle M_1M_2M_3$的面积S(图 5)，假如它的顶点的坐标分别是(x_1, y_1), (x_2, y_2), (x_3, y_3). 我们把已知三角形的面积看做是两底为y_1, y_3和y_3, y_2的两个梯形的面积之和减去两底为y_1, y_2的梯形的面积，而且把乘积$-(y_1+y_2)(x_2-x_1)$改写成$(y_1+y_2)\times(x_1-x_2)$，我们就得到

$$S=\frac{1}{2}[(y_1+y_2)(x_1-x_2)+(y_2+y_3)(x_2-x_3)+(y_3+y_1)(x_3-x_1)].$$

在以上几个问题里还需要验证，当有一个或几个坐标或者坐标之差是负数时，得到的公式并无任何改变而依然正确. 这种验证是并不难做的.

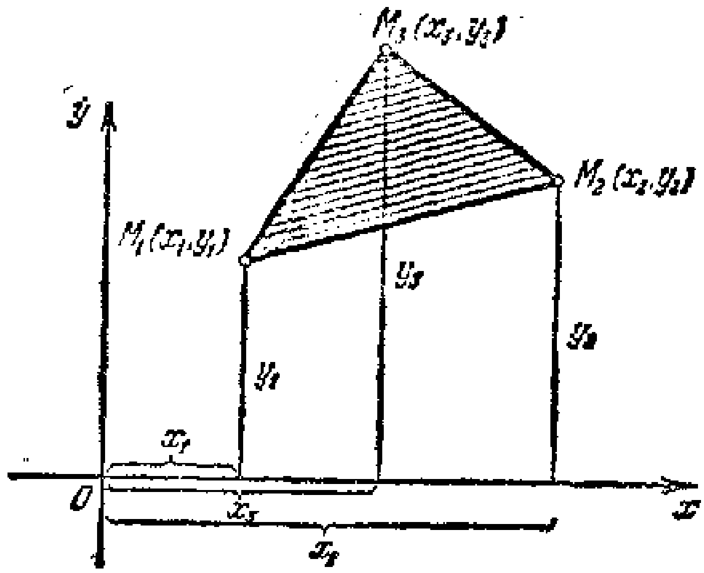


图 5

求两条曲线的交点 运用第二个基本观念(即认为方程$F(x, y)=0$表示曲线的观点)，求两条曲线的交点是特别简单的. 为了求两条曲线的交点的坐标，显然只需要联立地解表示这两条曲线的方程. 从这两个方程的普通的解得到的一对数x, y就给

出一个点，它的坐标既满足第一个方程，也满足第二个方程，即该点既在第一条曲线上，也在第二条曲线上，所以这点就是这两条曲线的交点.

我们看到，运用解析几何的方法来解几何问题在实用上极为方便，这特别是因为所有的解都直接以现成的数的形状而得到. 这样的几何学，这样的科学，怎能不恰恰合于那个时代呢？

§4. 由一次和二次方程所表示的曲线的研究

一次方程 在利用笛卡儿的第二个观念时，首先就会考虑什么样的曲线对应于一次方程，即方程

$$Ax+By+C=0, \tag{1}$$

这里 A, B, C 是一些数值系数，而且 A 和 B 不同时等于零. 可以证明，在平面上与这种方程对应的总是直线.

我们来证明，方程(1)总表示直线，而且反过来，对应于平面上每一条直线的是一个完全确定的方程(1). 实际上，例如设 $B\neq 0$. 那末方程(1)就可以解出

$$y=kx+l,$$

这里 $k=-\frac{A}{B}$; $l=-\frac{C}{B}$.

我们先来讨论方程 $y=kx$. 它显然表示这样一条直线，这条直线通过坐标原点而且它与 x 轴所成的角 φ 的正切 $\operatorname{tg}\varphi$ 就是 k (图 6). 实际上，这个方程可以改写成 $\frac{y}{x}=k$，而且上述直线的所有的点 (x, y) 都满足这个方程，至于不在这条直线上的任何点 (x, y)，因为对于它们说 $\frac{y}{x}$ 不是大于 k，就是小于 k，都不能满足这个方程. 这时，如果 $\operatorname{tg}\varphi>0$，则对于这种直线上的点说，$x$ 和 y 或者都是正的，或者都是负的，又如果 $\operatorname{tg}\varphi<0$，则 x 和 y 的符号是相反的.

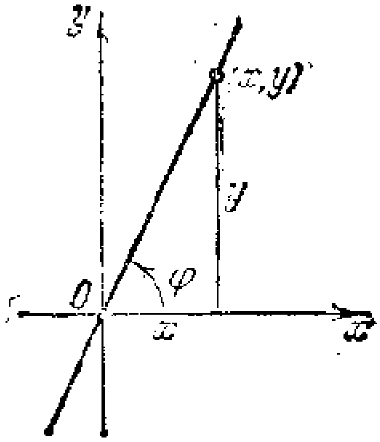


图 6

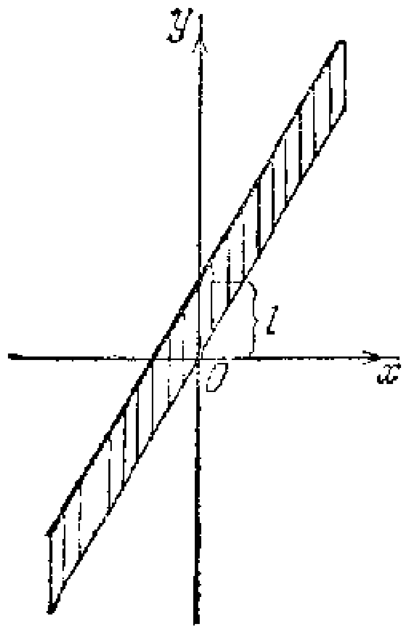


图 7

总之，方程 $y=kx$ 表示通过坐标原点的直线，因此，方程 $y=kx+l$ 也代表一条直线，那就是由前一条直线经过平行移动．使它的所有点的纵坐标都增加同一个 l 而得到的直线(图 7).

前面讨论过的几个公式：分线段成已知比值的点的坐标，两个已知点之间的距离，三角形的面积，以及现在关于直线方程的知识，已经能够用来解决很多问题了．

通过一个或两个已知点的直线的方程 设 M_1 是有坐标 x_1 和 y_1 的某个点，k 是某个已知数．方程 $y=kx+l$ 表示一条直线，它与 Ox 轴组成的角的正切等于 k，而且它从 Oy 轴截下的线段等于 l．我们取 l 使这条直线通过点(x_1, y_1)．为此，点 M_1 的坐标必须满足已知方程，即必须有 $y_1=kx_1+l$，于是 $l=y_1-kx_1$.

用这个 l 代入，我们就得到通过已知点(x_1, y_1)而且与 Ox 轴组成正切等于 k 的角的直线的方程(图 8)．它就是 $y=kx+y_1-kx^1$ 或者

$$y-y_1=k(x-x_1).$$

例 设直线和 Ox 轴之间的角等于 $45°$，而点 M 有坐标$(3,7)$，那末对应直线的方程(因为 $\operatorname{tg} 45°=1$)就是

$$y-7=1(x-3) \quad 或者 \quad x-y+4=0.$$

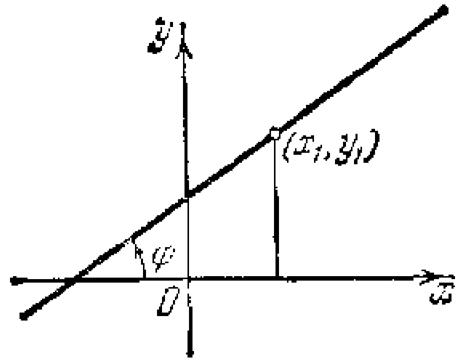


图 8

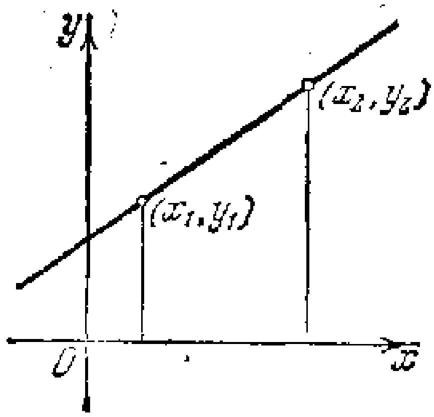


图 9

如果要求直线在通过点(x_1, y_1)的同时，还通过点(x_2, y_2)，那末我们就得到加于 k 上的条件 $y_2-y_1=k(x_2-x_1)$. 由此求得 k 并且代入前面的方程，我们就得到通过两个已知点的直线的方程(图 9)：

$$\frac{x-x_1}{x_2-x_1}=\frac{y-y_1}{y_2-y_1}.$$

笛卡儿关于二次方程的结果 笛卡儿还研究了具两个变数的二次方程表示平面上什么曲线的问题，这种二次方程的一般形状是

$$Ax^2+Bxy+Cy^2+Dx+Ey+F=0,$$

而且他还指出，这种方程一般地表示椭圆、双曲线或者抛物线——这些都是古代数学家所早已知道的曲线.

以上所说的都是笛卡儿的最重要的成就. 然而笛卡儿的书上远远不止限于这一些；笛卡儿还研究了一系列有趣的几何轨迹的方程，讨论了关于变换代数方程的定理，未加证明地导出了他的著名的符号法则，那是用来求出只有实根的方程的正根个数的(参看第四章，§4)，最后，他还提出了利用抛物线 $y=x^2$ 与圆的交点来求三次和四次方程的实根的卓越方法.

§5. 解三次和四次代数方程的笛卡儿方法

把三次和四次方程变换成没有 x^3 项的四次方程 我们来证

明，解三次和四次的任意方程可以化成解下列形状的方程：

$$x^4+px^2+qx+r=0. \tag{2}$$

设给了三次方程 $z^3+az^2+bz+c=0$. 令 $z=x-\frac{a}{3}$，我们得到 $\left(x-\frac{a}{3}\right)^3+a\left(x-\frac{a}{3}\right)^2+b\left(x-\frac{a}{3}\right)+c=0$. 去括弧后 x^2 项互相抵消，我们得到方程 $x^3+px+q=0$. 这个方程乘上 x，使它增添一个根 $x_4=0$，我们就把它化成(2)式，其中 $r=0$.

对于四次方程 $z^4+az^3+bz^2+cz+d=0$ 来说，只要令 $z=x-\frac{a}{4}$，就可以化成(2)式. 因此，解任何三次和四次方程都可以化成解形状(2)的方程.

用圆与抛物线 $y=x^2$ 的交点来解三次和四次方程 我们先建立中心为 (a, b) 和半径为 R 的圆的方程. 如果 (x, y) 是某个点，则它到点 (a, b) 的距离的平方等于 $(x-a)^2+(y-b)^2$（参看 §3）. 因此所讨论的圆的方程是

$$(x-a)^2+(y-b)^2=R^2.$$

现在我们再来求出这个圆与抛物线 $y=x^2$ 的交点. 根据 §3 中所说的，为此必须将这个圆的方程和抛物线的方程联立并求解：

$$x^2+y^2-2ax-2by+a^2+b^2-R^2=0,$$

$$y=x^2.$$

从第二个方程把 y 代入第一个方程，我们就得到关于 x 的四次方程

$$x^2+x^4-2ax-2bx^2+a^2+b^2-R^2=0,$$

或者方程 $\quad x^4+(1-2b)x^2-2ax+a^2+b^2-R^2=0.$

如果这样地选取 a, b 和 R^2，使得

$$1-2b=p,\ -2a=q,\ a^2+b^2-R^2=r,$$

则得到的正是方程(2). 为此必须取

$$a=-\frac{q}{2},\ b=\frac{1-p}{2},\ R^2=\frac{q^2}{4}+\frac{(1-p)^2}{4}-r. \tag{3}$$

在(3)中的末一个式子里，一般地说，R^2 可能是负的. 然而在方程(2)至少有一个实根 x_1 的情形，下列等式成立：

$$x_1^4+(1-2b)x_1^2-2ax_1+a^2+b^2-R^2=0, \qquad (4)$$

把 x_1^2 记做 y_1，等式(4)可以改写成

$$x_1^2+y_1^2-2ax_1-2by_1+a^2+b^2-R^2=0,$$

或者
$$(x_1-a)^2+(y_1-b)^2=R^2.$$

因此，在方程(2)有实根的情形，数 $R^2=\dfrac{(1-p)^2+q^2}{4}-r$ 是正数，方程

$$(x-a^2)+(y-b)^2=R^2$$

是圆的方程，而且方程(2)的全部实根都是抛物线 $y=x^2$ 与这个圆的交点的横坐标（在 $r=0$ 的情形，$R^2=a^2+b^2$，这个圆通过坐标原点）。

这样一来，如果给定了方程(2)的系数 p, q, r，而且按公式(3)求出了 a, b, R^2，那末当 $R^2<0$ 时，方程(2)显然没有实根．而假如 $R^2\geqslant 0$，则中心为 (a, b) 和半径为 R 的圆与固定地画着的抛物线 $y=x^2$ 的交点的横坐标，就给出方程(2)的全部实根，并且在 $R^2>0$ 的情形，得到的圆可以不与抛物线相交，即方程(2)可以没有实根．

例　设给了四次方程

$$x^4-4x^2+x+\frac{5}{2}=0.$$

于是我们有

$$a=-\frac{1}{2},$$

$$b=\frac{5}{2}=2\frac{1}{2},$$

$$R=\sqrt{\frac{1}{4}+\frac{25}{4}-\frac{5}{2}}=2.$$

在图 10 上画着对应的圆和所讨论的方程的根 x_1, x_2, x_3, x_4．

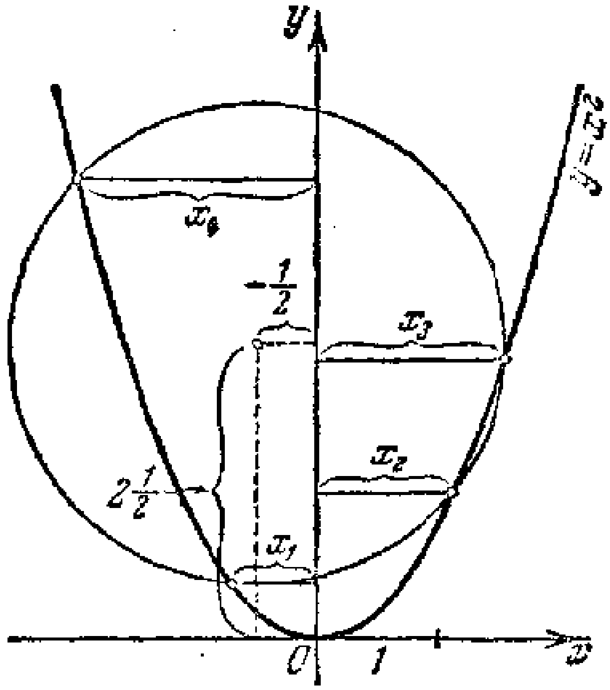


图　10

在§2, 3, 4, 5节里包含的是笛卡儿书中最主要的部分，但是我们用的是简短的而且更

接近于现代的叙述.

从笛卡儿的时代到今天，解析几何经历了有助于数学各部门的漫长的发展道路. 我们在本章以后各节里将要试着追循这路程上最重要的各个阶段.

首先应该指出，无穷小分析的发明者就已运用了笛卡儿的方法. 不管是关于曲线的切线或法线(切线在切点处的垂直线)的问题，还是关于函数的极大或极小的问题，假如几何地来讨论它的话，或者是关于曲线在其一点处的曲率半径的问题等等，首先都要用笛卡儿的方法来讨论这曲线的方程，然后还得求出切线的方程，法线的方程，等等. 所以无穷小分析、微分法和积分法没有解析几何的预先发展是难以想象的.

§6. 牛顿关于直径的普遍理论

第一个把解析几何向前推进了一步的是牛顿. 在1704年，他讨论了三次曲线(即由带两个未知数的三次代数方程表示的曲线)的理论，在这同一个工作中，牛顿顺便得出优美的关于“直径”(对应于已知方向的割线)的普遍理论. 下面就要来说明它.

设给了一条 n 次曲线，即由带两个未知数的 n 次代数方程表示的曲线，那末与它相交的任意直线，一般说来与它有 n 个公共点. 设 M 是割线上这样的点，它是割线与所说 n 次曲线的这些交点的“重心”，即分布在这些交点上的 n 个彼此相等的质点的“重心”. 可以证明，如果取所有彼此平行的割线，而且对于每条割线都考虑这样的重心 M，则所有这些点 M 处在一条直线上. 牛顿把这条直线叫做 n 次曲线的“直径”(对应于割线的已知方向). 利用解析几何来证明这个定理一点也不困难，所以我们要来做一下.

设给了一条 n 次曲线和它的一些彼此平行的割线. 我们选取这样的坐标轴，使得这些割线平行于 Ox 轴(图 11). 于是它们的方程就有形状 $y=l$，这里 l 是随割线的不同而有差别的常数. 设 $F(x, y)=0$ 是在这样的坐标系里表示所说 n 次曲线的方程. 容

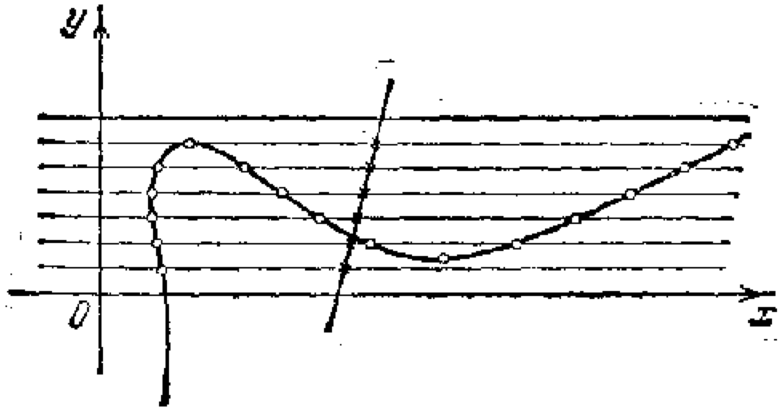


图 11

易证明，当从一个直角坐标系过渡到另一个直角坐标系时，虽然曲线的方程改变了，但是曲线的次数并不改变(这将在§8里证明)．所以 $F(x, y)$ 还是 n 次多项式．为了求出所说曲线与割线 $y=l$ 的交点的横坐标，必须联立地解方程 $F(x, y)=0$ 和 $y=l$，结果得到的一般说来是 x 的 n 次方程

$$F(x, l)=0, \tag{5}$$

从这方程可以求得横坐标 $x_1, x_2, \cdots, x_n$．按照重心的定义，n 个交点的重心的横坐标 x_0 等于

$$x_0=\frac{x_1+x_2+\cdots+x_n}{n}.$$

但是，从代数方程的理论知道，方程的诸根之和 $x_1+x_2+\cdots+x_n$，等于未知数的 $n-1$ 次项的系数取相反的符号再被 n 次项的系数除．而因为 x 和 y 的指数之和在 $F(x, y)$ 的每一项都等于或小于 n，所以具有 x^n 的项根本不包含 y 而有形状 Ax^n，这里 A 是常数，而具有 x^{n-1} 的项即使包含 y，也不会高于一次，即有形状 $x^{n-1}(By+C)$．因此，x^n 的系数是 A，而 x^{n-1} 的系数是 $Bl+C$，于是对于已知 l，我们就有

$$x_0=-\frac{Bl+C}{nA}.$$

但是割线平行于 Ox 轴，对于它的全部点都有 $y=l$，因此割线与所说 n 次曲线的交点的重心的纵坐标也等于 l；这样一来，我们最终地得到 $nAx_0+By_0+C=0$，即所有这些割线的全部被考虑的重心

的坐标 x_c, y_c 都满足一次方程，因此这些重心处在一条直线上.

在 $F(x, y)$ 里不出现 x^n 项时，可以作类似的研讨.

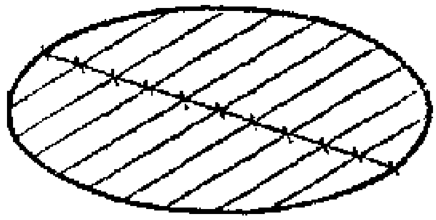


图 12

在二次曲线($n=2$)的情形，两个点的重心就是这两个点之间的中点，所以得到的结论是：二次曲线的平行弦中点的几何轨迹是直线(图 12)，不论是对于椭圆，还是对于双曲线和抛物线，这是早在古代就已经知道了的. 然而即使对于这些特别情形，要用几何的论证来证明这一点也是非常困难的，可是这个古代所不知道的新的普遍定理，现在却非常简单地证明了.

这样的例子显示出解析几何的力量.

§7. 椭圆、双曲线和抛物线

在这一节和下一节我们要来讨论二次曲线. 首先研究一般的二次方程，然后讨论它的一些最简单的特例.

以坐标原点为中心的圆的方程 我们先讨论方程

$$x^2+y^2=a^2.$$

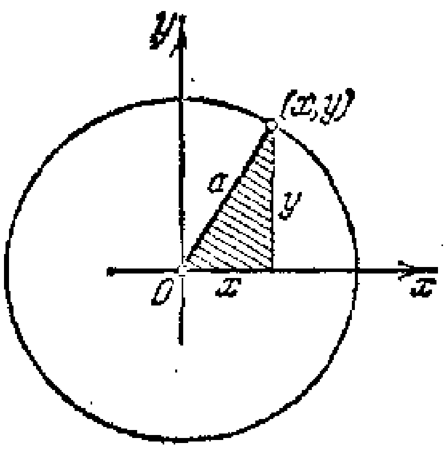


图 13

它显然表示以坐标原点为中心而且有半径 a 的圆，这从对加了斜线的直角三角形(图 13)引用毕达哥拉斯定理而得出，因为不管在这个圆上取怎样的一个点 (x, y)，它的坐标 x 和 y 满足这个方程，而且反之，如果一个点的坐标 x, y 满足这个方程，则它就属于这个圆，这就是说，这个圆是平面上满足这个方程的所有点的集合.

椭圆的方程和椭圆的焦点性质 设给了两个点 F_1 和 F_2，它们之间的距离等于 $2c$，我们来求平面上到点 F_1 和 F_2 的距离之

和等于一个常数 $2a$(这里 a 自然要大于 c)的所有点 M 的几何轨迹. 这样的曲线叫做椭圆, 点 F_1 和 F_2 叫做椭圆的焦点.

我们这样地选取直角坐标系, 使得点 F_1 和 F_2 都在 Ox 轴上, 而且坐标原点正是它们之间的中点. 于是点 F_1 和 F_2 的坐标就是 $(c, 0)$ 和 $(-c, 0)$. 我们在所研究的几何轨迹上取任意点 M, 设它有坐标 (x, y), 然后写出从点 M 到点 F_1 和 F_2 的距离之和等于 $2a$, 即

$$\sqrt{(x-c)^2+(y-0)^2}+\sqrt{(x+c)^2+(y-0)^2}=2a. \qquad (6)$$

所说几何轨迹上的任意点的坐标 (x, y) 都满足这个方程, 显然反过来也对, 那就是说, 坐标满足方程(6)的任意点都属于这个几何轨迹. 因此, 方程(6)是所说几何轨迹的方程. 还需要把它简化.

两边平方, 我们得到

$$x^2-2cx+c^2+y^2+2\sqrt{(x^2-2cx+c^2+y^2)(x^2+2cx+c^2+y^2)}$$
$$+x^2+2cx+c^2+y^2=4a^2,$$

或者经过简化以后,

$$x^2+y^2+c^2-2a^2=-\sqrt{(x^2+y^2+c^2)^2-4c^2x^2}.$$

两边再平方, 我们得到

$$(x^2+y^2+c^2)^2-4a^2(x^2+y^2+c^2)+4a^4=(x^2+y^2+c^2)^2-4c^2x^2,$$

或者经过简化以后,

$$(a^2-c^2)x^2+a^2y^2=(a^2-c^2)a^2.$$

在这里令 $a^2-c^2=b^2$ (因为 $a>c$, 这是可以做到的), 我们得到 $b^2x^2+a^2y^2=a^2b^2$, 两边都除以 a^2b^2, 我们就有

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1. \qquad (7)$$

这样一来, 所说几何轨迹上的任意点 M 的坐标 (x, y) 满足方程(7).

可以证明, 反过来, 如果一个点的坐标满足方程(7), 则它也满足方程(6). 因此, 方程(7)是这个几何轨迹的方程, 即是椭圆的方程(图 14).

以上所引的论证是寻求由其若干几何性质给定的曲线方程的

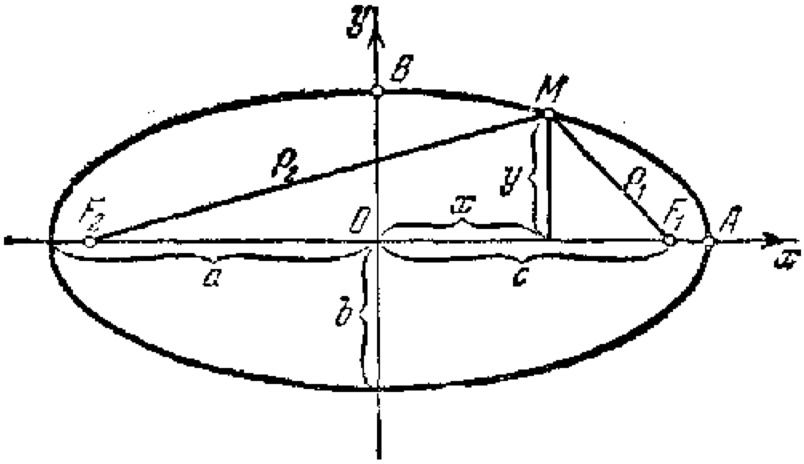


图 14

典型例子.

以椭圆的任意点到两个已知点的距离之和是常数这个性质为基础，我们有利用丝线来画椭圆的熟知方法(图 15).

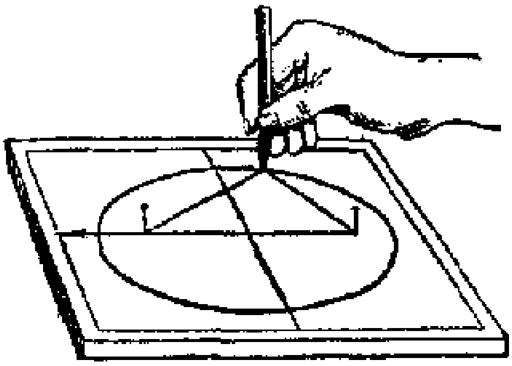


图 15

附注 为了定义什么曲线叫做椭圆，可以不取这里所说的椭圆的焦点性质，而取任何别的能刻画它的几何性质. 例如椭圆是圆向它的直径作"均匀压缩"的结果这一性质(参看第 242 页)，或者某个别的性质.

在方程(7)上令 $y=0$, 我们得到 $x=\pm a$, 即 a 是线段 OA 的长度(图 14)，这线段叫做椭圆的长半轴. 类似地，令 $x=0$, 我们得到 $y=\pm b$, 即 b 是线段 OB 的长度，这线段叫做椭圆的短半轴.

数 $\varepsilon=\frac{c}{a}$ 叫做椭圆的离心率，并且因为 $c=\sqrt{a^2-b^2}<a$, 所以椭圆的离心率小于 1. 在圆的情形 $c=0$, 因此 $\varepsilon=0$; 这时两个焦点合成一个点——圆的中心(因为 $OF_1=OF_2=0$)，但是前面用丝线画椭圆的方法仍然能用.

行星的运动规律 刻普勒在研究了钛何·勃拉赫关于火星在天空中运动的大量观察材料以后，发现行星沿着椭圆轨道而绕着

太阳运行，这时太阳处在这椭圆的一个焦点上（在行星绕太阳的运动中，另一个焦点不起任何作用）（图 16），并且焦半径 ρ 在同样的时间内扫过同样面积的扇形[1]，而牛顿则指明，这种运动的必要性可以数学地从惯性定律、加速度与作用力成正比的定律和万有引力定律推得.

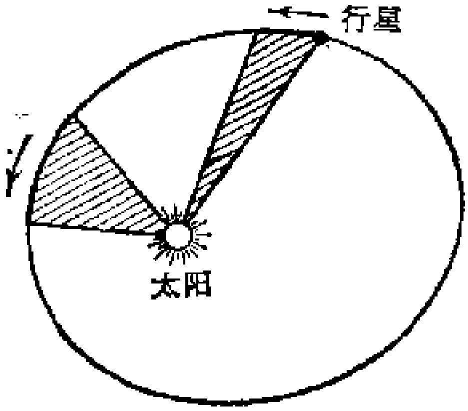


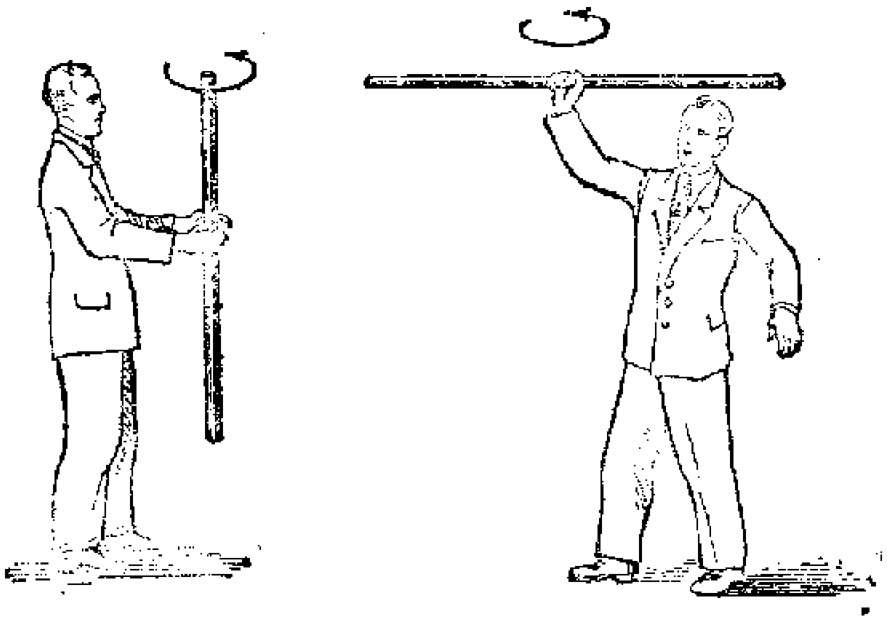
图 16

惯性椭圆 作为椭圆应用在技术问题上的例子，我们来讨论薄板的所谓惯性椭圆. 设有着一块同样厚薄而且由均匀的物质作成的薄板，例如某种形状的一块洋铁板. 我们使它绕着在板平面上的一条轴旋转. 大家知道，沿直线运动的物体，对于这直线运动说有一个与其质量成正比的惯性（不依赖于物体的形状和其质量的分布）. 与这相仿，绕轴旋转的物体，例如飞轮，对于这转动说也有惯性. 然而在旋转的情形里，惯性不仅与旋转物体的质量成正比，而且还依赖于这物体的质量对于旋转轴的分布，因为质量距离旋转轴愈远，对于旋转的惯性就愈大. 譬如说，一根棒非常容易使它立刻绕纵轴作迅速的旋转（图 17 a）. 而如果要它立刻绕垂直于其长的轴作迅速的旋转，则即使令旋转轴通过它的中点，当这根棒不是很轻时，也要费去很大的力量（图 17 b）.

可以证明，物体绕某条轴旋转的惯性，即所谓物体关于这条轴的“惯性矩”，等于 $\sum r_i^2 m_i$. （$\sum r_i^2 m_i$ 是和式 $r_1^2 m_1 + r_2^2 m_2 + \cdots + r_n^2 m_n$ 的缩写，并且认为物质分割成按其量说是非常小的微粒，而 m_i 是微粒的质量，r_i 是微粒到旋转轴的距离，而且对物体的全部微粒求总和.）

回到我们的薄板. 设 O 是这薄板的任意点（图 18）. 我们来讨论这薄板对于通过点 O 而且处在薄板平面上的轴 u 的惯性矩 J_u. 为此我们取 O 作为直角坐标系的原点，再在薄板平面上随便选取

1) 行星轨道的离心率极小，因而行星的轨道几乎是一个圆.



a　　　b

图 17

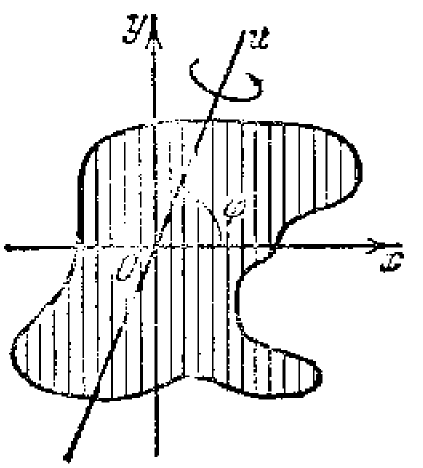


图 18

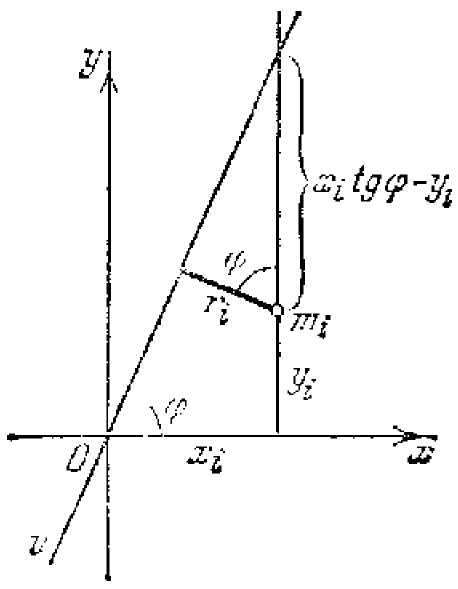


图 19

Ox 轴和 Oy 轴，于是旋转轴 u 就可以用它与 Ox 轴的交角 φ 来刻画．容易看出(图 19)，

$$r_i=|(x_i \operatorname{tg}\varphi-y_i)\cos\varphi|=|x_i\sin\varphi-y_i\cos\varphi|.$$

由此

$$\begin{aligned}\sum r_i^2 m_i&=\sum(x_i^2\sin^2\varphi-2x_iy_i\sin\varphi\cos\varphi+y_i^2\cos^2\varphi)m_i\\&=\sin^2\varphi\sum x_i^2m_i-2\sin\varphi\cos\varphi\sum x_iy_im_i+\cos^2\varphi\sum y_i^2m_i.\end{aligned}$$

量 $\sin^2\varphi$，$2\sin\varphi\cos\varphi$ 和 $\cos^2\varphi$ 可以拿出和号之外，因为这些量对

于轴 u 说都是常量. 现在我们记

$$\sum x_i^2 m_i = A, \quad -\sum x_i y_i m_i = B, \quad \sum y_i^2 m_i = C.$$

量 A, B 和 C 不依赖于轴 u 的选取, 而只依赖于薄板的形状和其质量的分布以及坐标轴 Ox 和 Oy 的一劳永逸的选择. 总之,

$$J_u = A\sin^2\varphi + 2B\sin\varphi\cos\varphi + C\cos^2\varphi.$$

我们来讨论在薄板平面上通过点 O 的所有可能的轴 u. 我们在每条轴上从点 O 起截取一个长度 ρ 的线段, 使 ρ 等于薄板对于这轴的惯性矩 J_u 的平方根的倒数, 即 $\rho = \frac{1}{\sqrt{J_u}}$. 于是我们得到

$$\frac{1}{\rho^2} = A\sin^2\varphi + 2B\sin\varphi\cos\varphi + C\cos^2\varphi.$$

但是 $$x = \rho\cos\varphi, \quad y = \rho\sin\varphi,$$

因此, 这个几何轨迹的方程有下列形状:

$$Cx^2 + 2Bxy + Ay^2 = 1.$$

得到的是一条二次曲线, 它显然是有界的和封闭的, 即是一个椭圆(图 20), 因为我们在以后将要证明, 所有其他的二次曲线, 不是无界的, 就是退化成一点.

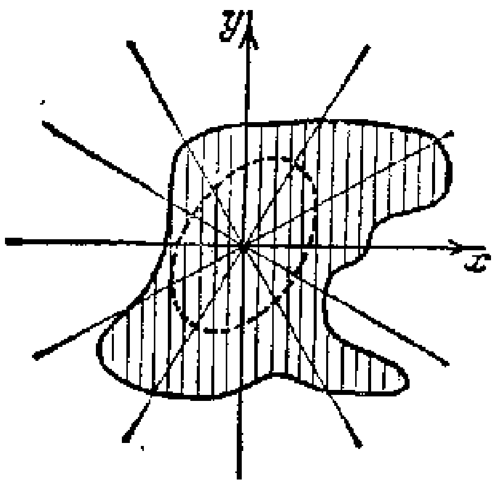


图 20

现在得到了一个值得注意的结果: 不管薄板的形状和大小如何, 也不管其质量如何分布, 这薄板对于处在薄板平面而且通过薄板的已知点 O 的各条轴的惯性矩(特别是与它们的平方根成反比的量 ρ), 刻画出一个椭圆. 这个椭圆叫做薄板对于点 O 的惯性椭圆. 如果点 O 是薄板的重心, 则这个椭圆叫做薄板的中心惯性椭圆.

惯性椭圆在力学中起着很大作用, 而且它在材料强度里有特别重要的应用. 在材料强度里证明, 假如我们有一根已知截面的横梁, 则它的抗弯强度与其截面对于通过截面重心而垂直于弯曲力方向的轴的惯性矩成正比. 让我们用例子来说明这一点. 假定

小河上的小桥是木板做的，这木板受到过桥行人的重量作用而弯曲。如果同一块木板(而不是较厚的木板)侧面朝上放，它几乎完全不会弯曲，这就是说，侧面朝上放的木板是所谓坚固的。这是因为木板的横截面是相当长的长方形，这截面对于处在其平面上，通过其中心而且垂直于其长边的轴的惯性矩(假如认为截面上质量是均匀分布的)，比对于平行于其长边的轴的惯性矩来得大。如果不是平放，也不是侧着放，而是斜着放，甚至如果取的不是木板，而是有任意截面的材料，例如铁轨，弯曲强度依然与截面对于处在其平面上而且通过其重心的轴的惯性矩成正比。因此，横梁经受弯曲的强度被其截面的惯性椭圆所刻画。

例如，对于普通的长方形截面的木板，这个椭圆有画在图 21 上的形状。这种横梁在 Oz 轴方向上经受弯曲的强度与 bh^3 成正比。

钢梁常常采取"⅂"形的截面；在图 22 上画着这种横梁的截面和惯性椭圆。它在 z 轴的方向具有经受弯曲的最大强度。例如在雪的负荷和特殊的重压下的屋梁，就可以利用它们，这时它们正好在这最适宜的方向上经受弯曲。

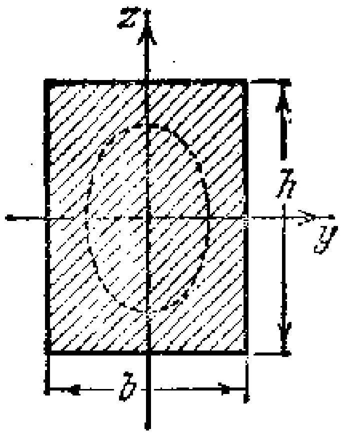


图 21

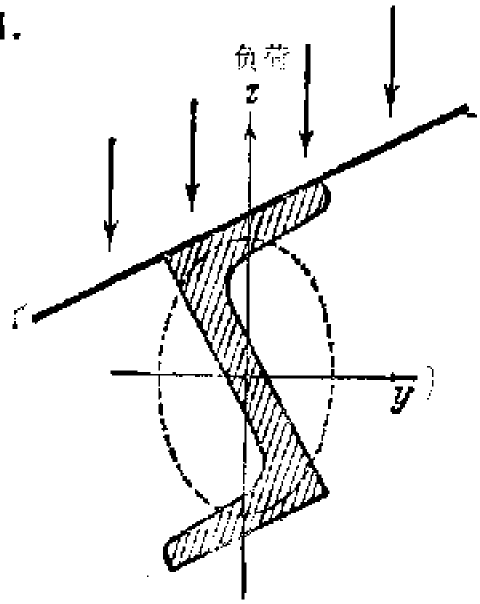


图 22

双曲线和它的焦点性质 现在我们来讨论方程

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1,$$

它表示的曲线叫做双曲线. 如果用 c 表示使 $c^2=a^2+b^2$ 的数, 则可以证明, 双曲线是到两个已知点 F_1 和 F_2 的距离之差是常量 $\rho_2-\rho_1=2a$ 的点的几何轨迹, 这里点 F_1 和 F_2 处在 Ox 轴上而且有横坐标 c 和 $-c$(图 23). 点 F_1 和 F_2 也叫做焦点.

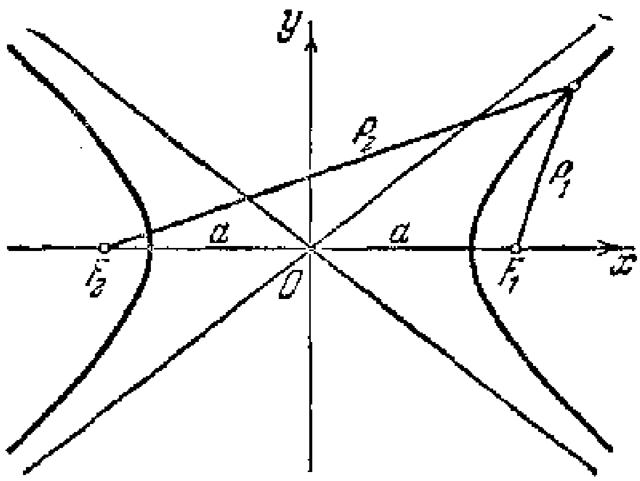


图 23

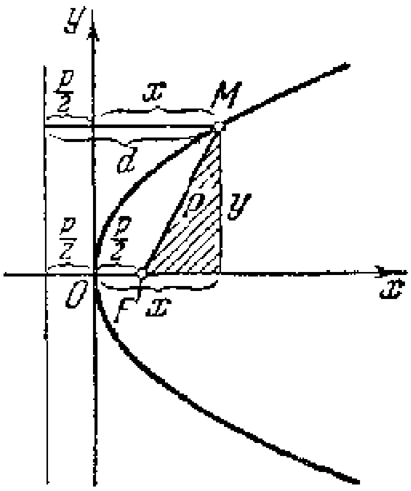


图 24

抛物线和它的准线 最后我们来讨论方程

$$y^2=2px,$$

我们把它所表示的曲线叫做抛物线. 处在 Ox 轴上而且有横坐标 $\frac{p}{2}$ 的点 F 叫做抛物线的焦点, 平行于 Oy 轴的直线 $y=-\frac{p}{2}$ 叫做抛物线的准线. 设 M 是抛物线上的一个点(图 24), 而且 ρ 是其焦半径 MF 的长度, d 是从 M 引到准线的垂直线的长度. 我们对于点 M 来计算 ρ 和 d. 从加斜线的三角形我们得到

$$\rho^2=\left(x-\frac{p}{2}\right)^2+y^2.$$

由于点 M 在抛物线上, 对于它有 $y^2=2px$, 因此

$$\rho^2=\left(x-\frac{p}{2}\right)^2+2px=\left(x+\frac{p}{2}\right)^2.$$

但是直接从图上看到 $d=x+\frac{p}{2}$, 所以 $\rho^2=d^2$, 即 $\rho=d$. 反面的论证表明, 如果有点满足条件 $\rho=d$, 则它就在所说的抛物线上. 因此, 抛物线是这种点的几何轨迹, 从其中每个点到已知点 F(所谓焦点)的距离 ρ, 等于它到已知直线(所谓准线)的距离 d.

抛物线切线的性质 我们来讨论抛物线切线的一个重要性质和它在光学上的应用．因为对于抛物线说有 $y^2=2px$，所以对于它说有 $2y\,dy=2p\,dx$，即导数(或者说切线对 Ox 轴的倾角 φ 的正切也一样)等于 $\frac{dy}{dx}=\operatorname{tg}\varphi=\frac{p}{y}$ (图 25)．

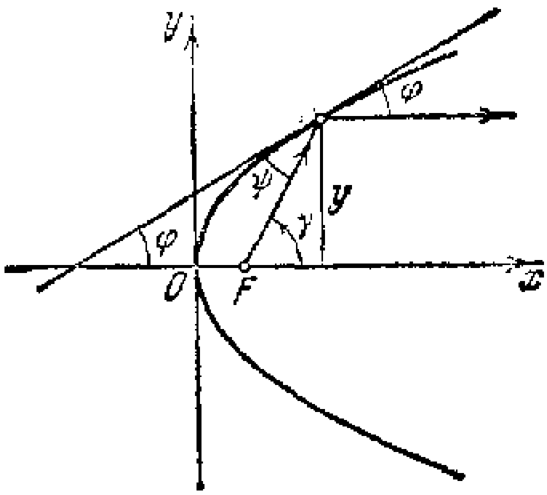


图 25

另一方面，直接从图上得出

$$\operatorname{tg}\gamma=\frac{y}{x-\frac{p}{2}}.$$

但是

$$\operatorname{tg}2\varphi=\frac{2\frac{p}{y}}{1-\frac{p^2}{y^2}}=\frac{2py}{y^2-p^2}=\frac{2py}{2px-p^2}=\frac{y}{x-\frac{p}{2}},$$

即 $\gamma=2\varphi$，又因为 $\gamma=\varphi+\psi$，所以 $\psi=\varphi$．于是从焦点 F 出发的射线，经过方向与切线方向重合的抛物线微弧的反射，根据入射角等于反射角的定律，反射成平行于 Ox 轴(即平行于抛物线的对称轴)的光线而远去．

牛顿根据抛物线的这个性质制造了反射望远镜．假如制造一个凹面镜，它的表面是一个所谓旋转抛物面，即由抛物线绕其轴旋转而得到的曲面，那末从天体某点发出的、严格地沿着镜"轴"的方向的所有光线，就被镜面聚集于一个点——镜面的焦点(图 26)．从天体上另外的点发出的、已经不完全平行于镜轴的光线，也几乎被聚集成靠近焦点的一个点．因此，在所谓焦平面上，即在通过镜面焦点而垂直于其轴的平面上，我们得到天体的一个倒像，并且这个像离开焦点愈远就愈模糊，因为只有严格地平行于镜轴的光线才聚集于一个点；而不完全平行于轴的光线并不完全聚集于一个点．得到的像可以在特制的显微镜(所谓透视望远镜)里观看，这

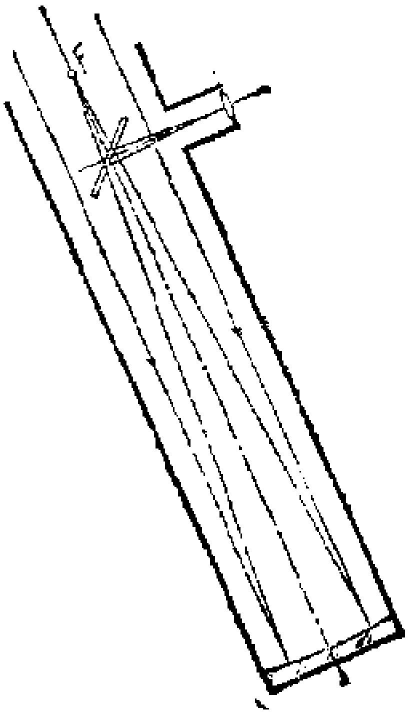


图 26

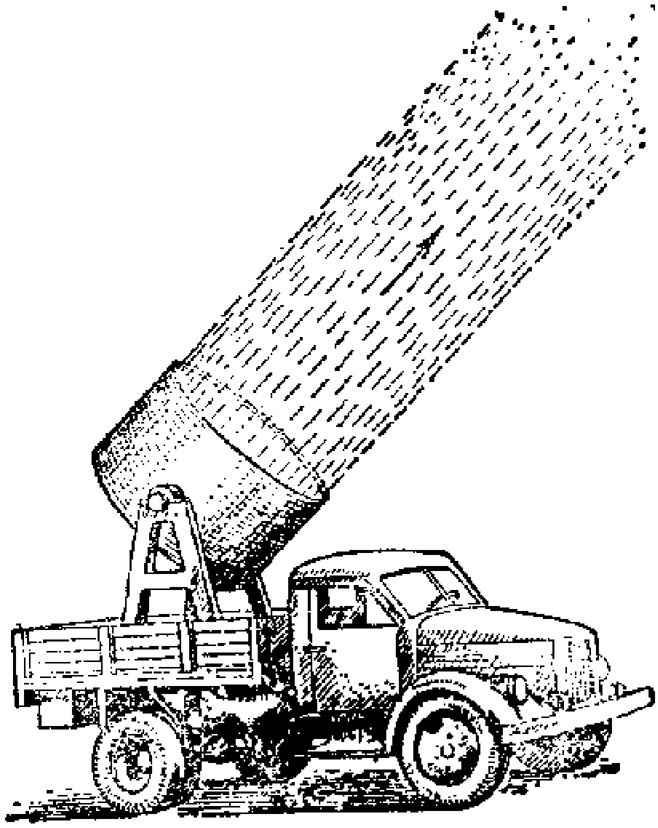


图 27

时可以直接看，也可以为了不使头部遮住天体，用一个不大的平面镜使光线转向来看，这平面镜放置在望远镜内的焦点附近（比焦点稍稍靠近凹面镜），与凹面镜轴作成 $45°$ 角.

根据抛物线的所说性质还可以制作探照灯(图 27). 在探照灯中，相反地，强有力的光源放在抛物面镜的焦点处，于是它的光线经过镜面的反射，就成为平行于镜轴的光线. 汽车的前灯也有同样的结构(图 28).

在椭圆的情形，容易证明，从它的一个焦点 F_1 出发的光线，

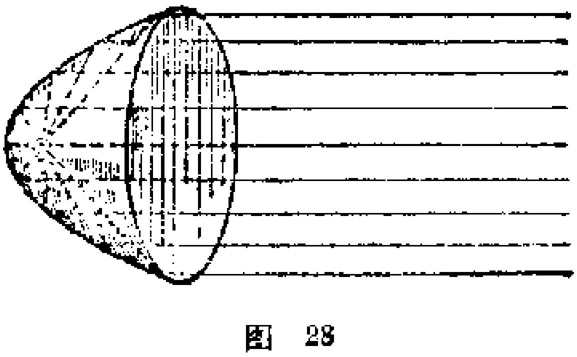


图 28

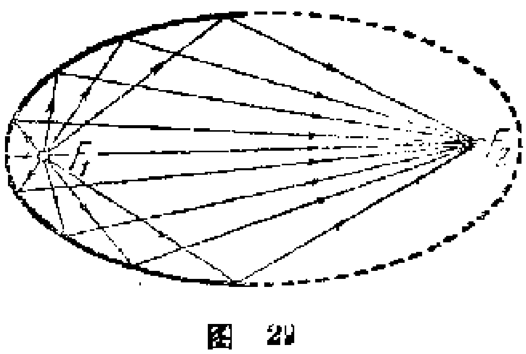


图 29

经过椭圆的反射，聚集于另一个焦点 F_2 处(图 29). 而在双曲线的情形，从它的一个焦点 F_1 出发的光线，经过它的反射后，成为好象从另一个焦点 F_2 出发的光线(图 30).

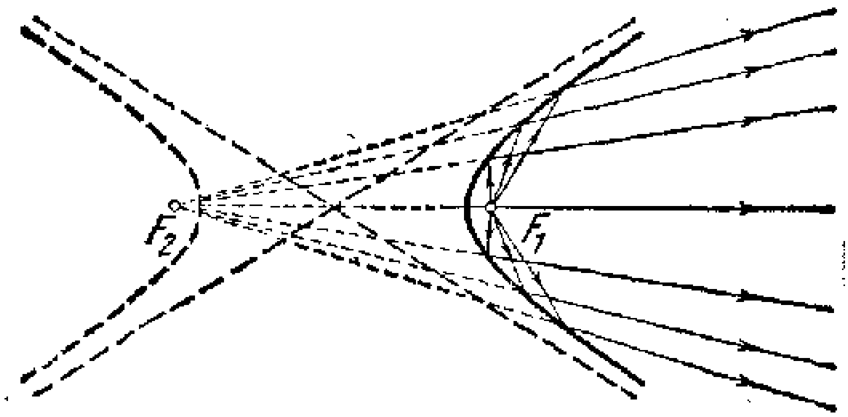


图 30

椭圆和双曲线的准线 与抛物线类似，椭圆和双曲线也有准线，并且有两条准线. 如果取焦点和与它同侧的准线，则对于椭圆的所有点都有 $\frac{\rho}{d}=\varepsilon$，这里 ε 是所说椭圆的离心率，在椭圆的情形它总小于 1，同时对于双曲线对应一支的所有点也都有 $\frac{\rho}{d}=\varepsilon$，这里 ε 是所说双曲线的离心率，并且在双曲线的情形它总大于 1.

因此，椭圆、抛物线和双曲线都是平面上所有这种点的几何轨迹，点到焦点的距离 ρ 对点到准线的距离 d 的比值是常数(图 31 和 32). 只是这个常数在椭圆时小于 1，在抛物线时等于 1，在双曲线时大于 1. 在这意义下，抛物线是从椭圆到双曲线的所谓"极

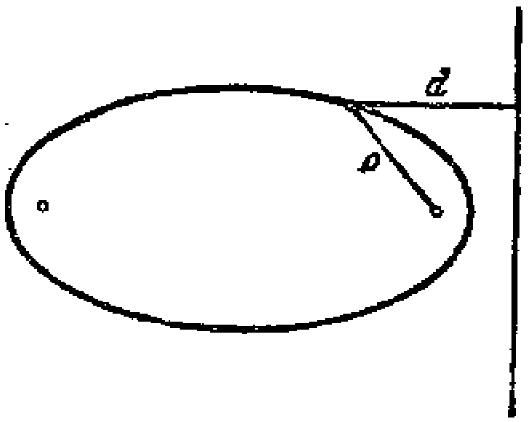


图 31

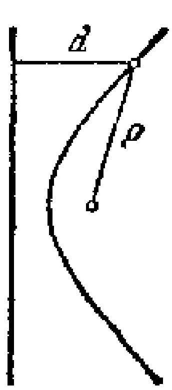


图 32

限"情形或者"过渡"情形.

圆锥截线 在古希腊时还曾详细地研究了正圆锥被平面所截而得到的曲线. 如果截平面与圆锥的轴作成的角φ是90°, 即如果平面垂直于轴, 则得到的截线是圆. 容易证明, 如果角φ小于90°, 但是大于圆锥的轴与其母线组成的角α, 则得到的是椭圆. 如

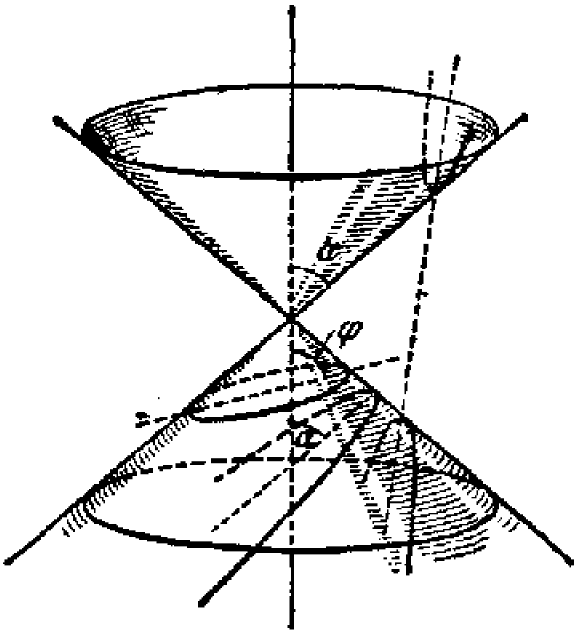


图 33

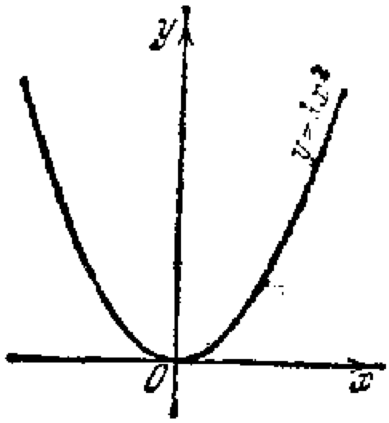


图 34

果角φ等于角α, 则得到的是抛物线. 如果角φ小于角α, 则得到的截线是双曲线(图33).

抛物线作为与平方成正比的图像和双曲线作为反比的图像 我们要来指出, 与平方成正比的函数

$$y=kx^2$$

的图像是抛物线(图34), 反比函数

$$y=\frac{k}{x} \quad 或 \quad xy=k$$

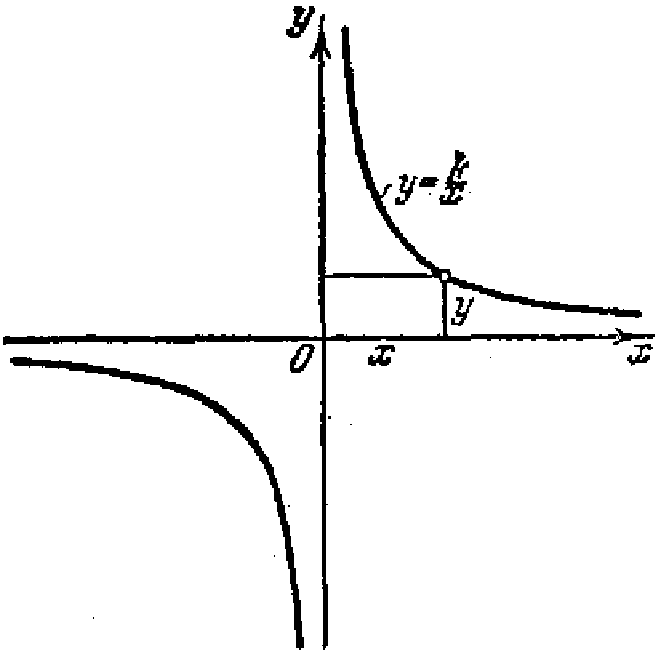


图 35

的图像是双曲线(图 35). 后者是很容易证实的. 双曲线在前面是被定义为由方程

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$$

所表示的曲线的.

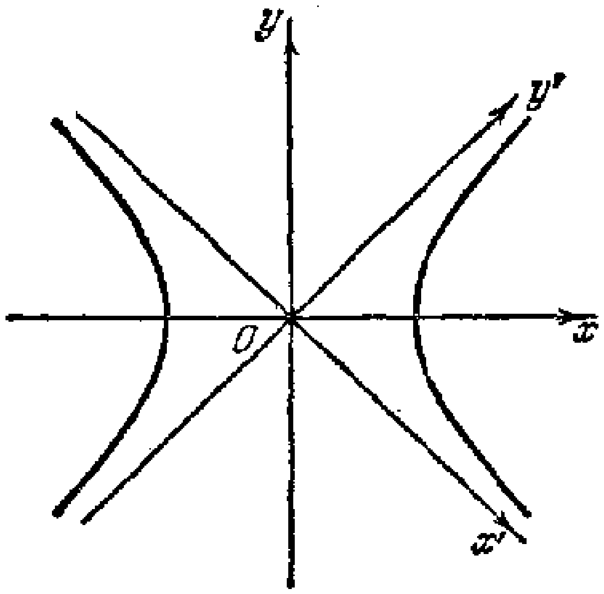


图 36

在特别情形, 当 $a=b$ 时, 这就是所谓"等边双曲线", 它在双曲线中起着与圆在椭圆中相同的作用. 在这种情形, 如果让坐标轴旋转 45°(图 36), 这个方程在新坐标系里就变成

$$x'y'=k.$$

刚才我们讨论了三种重要的二次曲线: 椭圆、双曲线和抛物线, 并且取作它们的定义的是用来表示它们的所谓"标准"方程:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \quad \frac{x^2}{a^2}-\frac{y^2}{b^2}=1 \quad 和 \quad y^2=2px.$$

现在我们要来研究带两个变数的一般的二次方程, 那就是说, 来研究它所表示的究竟是哪样一些曲线的问题.

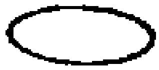
§8. 把一般的二次方程化成标准形状

欧拉所作的解析几何的第一个系统的叙述 1748 年在欧拉的«分析引论»里出现了解析几何发展中的重要的一步, 在该书第二卷里, 在其他事物之间, 先给出的是归于函数论和分析其他部分的东西, 但是也叙述了平面上的解析几何, 附有二次曲线的详细的研究, 非常接近于今天解析几何教科书中的叙述, 而且还研究了高阶曲线. 这是在现代字义下的第一本解析几何教程.

把方程化成标准形状的观念　二次方程[1)]

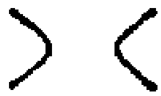
$$Ax^2+2Bxy+Cy^2+2Dx+2Ey+F=0$$

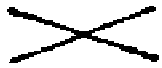
包含着六项，而不像前面所讨论的椭圆、双曲线和抛物线的标准方程那样只包含三项或两项．这并非由于这种方程表示更复杂的曲线，而是由于写出这方程时所在的坐标系对它并不是最适宜的．可以证明，如果适当地选用笛卡儿直角坐标系，则带两个变数的二次方程总可以化成下列标准形状中的一个：

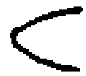
1) $\frac{x^2}{a^2}+\frac{y^2}{b^2}-1=0$，  椭圆

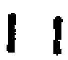
2) $\frac{x^2}{a^2}+\frac{y^2}{b^2}+1=0$，  虚椭圆

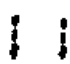
3) $\frac{x^2}{a^2}+\frac{y^2}{b^2}=0$， 点(一对相交于实点的虚直线)

4) $\frac{x^2}{a^2}-\frac{y^2}{b^2}-1=0$，  双曲线

5) $\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$，  一对相交的直线

6) $y^2-2px=0$，  抛物线

7) $x^2-a^2=0$，  一对平行的直线

8) $x^2+a^2=0$，  一对虚的平行直线

9) $x^2=0$，  一对重合的直线

这里 a, b, p 都不等于零．

在列举的标准形状中，方程(1)，(4)和(6)我们早已知道了；它们就是椭圆、双曲线和抛物线的标准方程．列举的方程中有两个不被任何实点所满足，那就是方程(2)和(8)．事实上，实数的平方总是正数或者零，因而在方程(2)的左边已经有两项之和 $\frac{x^2}{a^2}+\frac{y^2}{b^2}$ 不是负数，除此而外还有一项 $+1$，因此左边不能成为零；同

1）xy, x, y 的系数我们不记做 B, D, E, 而记做 $2B$, $2D$, $2E$, 是为了简化后面的公式的写法．

样地，在方程(8)中数 a^2 不是负数，而 a^2 则是正数. 从同一些想法知道，方程(3)只被一个点(0, 0)——坐标原点所满足. 方程(5)可以写成：

$$\left(\frac{x}{a}-\frac{y}{b}\right)\left(\frac{x}{a}+\frac{y}{b}\right)=0,$$

于是就知道，满足它的是而且只是使一次式 $\frac{x}{a}-\frac{y}{b}$ 或者 $\frac{x}{a}+\frac{y}{b}$ 等于零的那些点，即它所表示的是由这样一对相交直线的全体组成的曲线. 方程(7)同样地给出 $(x-a)(x+a)=0$，即对应的曲线是一对平行的直线 $x=a$ 和 $x=-a$. 最后，曲线(9)是曲线(7)当 $a=0$ 时的特别(极限)情形，即是一对重合的直线.

坐标变换的公式 为了得到关于二次曲线可能的类型的所说重要结果，我们必须先引出在坐标系改变时点的直角坐标如何改变的公式.

设 x, y 是某个点 M 对于坐标轴 Oxy 的坐标. 我们把这些坐标轴平行于自己地移到新位置 $O'x'y'$，而且新原点 O' 对于旧坐标轴的坐标是 ξ 和 η. 明显地(图 37)，点 M 对于新坐标轴的坐标 x' 及 y' 与它对于旧坐标轴的坐标 x 及 y 的关系由下列公式表示：

$$x=x'+\xi,$$
$$y=y'+\eta,$$

这就是所谓平行移轴的公式. 又如果我们让坐标轴 Oxy 绕原点向反时针方向旋转一个角 φ，则容易得知(图 38)，把由新坐标线

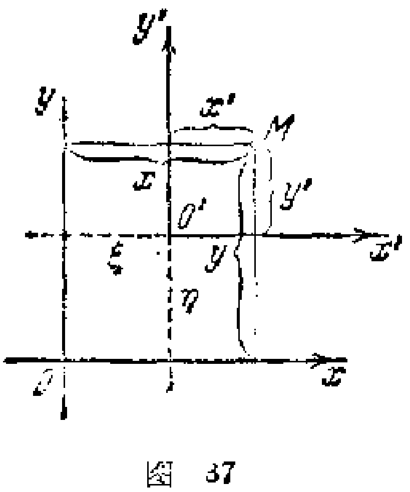


图 37

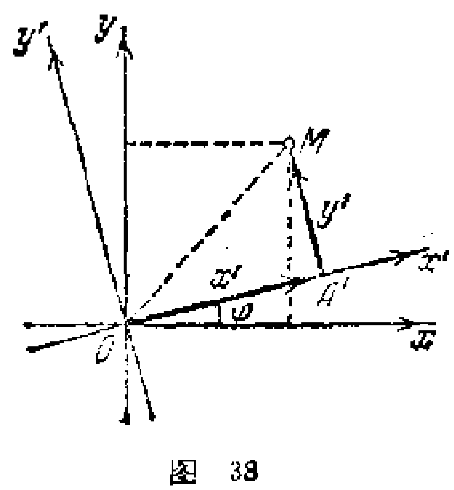


图 38

段 x' 及 y' 组成的折线 $OA'M$ 投射到 Ox 轴和 Oy 轴上，我们得到

$$x=x'\cos\varphi-y'\sin\varphi,$$
$$y=x'\sin\varphi+y'\cos\varphi,$$

这是旋转直角坐标系的坐标变换的公式[1]。

如果给了某条曲线对于坐标轴 Oxy 的方程 $F(x, y)=0$，而要写出这同一条曲线的经过变换的方程，即这同一条曲线对于新坐标轴 $O'x'y'$ 的方程，那末只要把方程 $F(x, y)=0$ 中的 x 和 y，用由变换公式所给的它们通过 x' 和 y' 的表达式代入就成，举例说，经过平行移轴，我们得到变换成的方程

$$F(x'+\xi,\ y'+\eta)=0,$$

而经过转轴则得到方程

$$F(x'\cos\varphi-y'\sin\varphi,\ x'\sin\varphi+y'\cos\varphi)=0.$$

我们注意到，在变换到新坐标轴时，方程的次数不会改变．实际上，因为变换公式是一次的，所以次数不能增高，但是次数也不能减低，因为否则经过反逆坐标变换，次数就要增高了（而后者也是一次的）．

把任意二次方程化成 9 种标准形状中的一种　现在让我们来证明，对于无论怎么样给定的带两个变数的二次方程，总可以先经过转轴，再经过平行移轴，使得在最终的坐标轴里，变换成的方程有形状(1)，(2)，…，(9)中的一个．

实际上，设给定的二次方程有形状

$$Ax^2+2Bxy+Cy^2+2Dx+2Ey+F=0. \tag{8}$$

让轴旋转一个角 φ，这个角是我们现在要来选取的．把方程(8)中的 x 和 y 用它们以新坐标表示的式子（按照转轴公式）代入，经过合并同类项，我们在变换成的方程

$$A'x'^2+2B'x'y'+C'y'^2+2D'x'+2E'y'+F'=0$$

里得到系数 $2B'$：

$$2B'=-2A\sin\varphi\cos\varphi+2B(\cos^2\varphi-\sin^2\varphi)+2C\sin\varphi\cos\varphi$$
$$=2B\cos2\varphi-(A-C)\sin2\varphi.$$

1）也叫做转轴的公式．——译者注

令这个系数等于零，我们得到

$$2B\cos 2\varphi=(A-C)\sin 2\varphi,$$

于是

$$\operatorname{ctg} 2\varphi=\frac{A-C}{2B}.$$

但是因为余切的变动范围是从 $-\infty$ 到 $+\infty$，所以总可以求得这样的角 φ，使得这个等式成立．当坐标轴旋转这样一个角后，在旋转后的坐标轴 $Ox'y'$ 里，我们得到在原来坐标轴里由方程(8)表示的曲线方程

$$A'x'^2+C'y'^2+2D'x'+2E'y'+F=0. \tag{9}$$

即它已经不再包含坐标乘积的项．(F 还是原来的，因为转轴公式不包含常数项.)

现在我们让已经旋转过的轴 $Ox'y'$ 平移到新位置 $O''x''y''$，而且设新原点 O'' 对于坐标轴 $Ox'y'$ 的坐标是 ξ', η'．我们的曲线在这些最终的坐标轴里的方程是

$$A'(x''+\xi')^2+C'(y''+\eta')^2+2D'(x''+\xi')$$
$$+2E'(y''+\eta')+F=0. \tag{10}$$

我们来证明，总可以如此地选取 ξ' 和 η'，即如此地让坐标轴 $Ox'y'$ 作平行于自己的移动，使得在坐标轴 $O''x''y''$ 里的最终的方程有标准形状(1)，(2)，…，(9)中的一个．

在方程(10)里脱去全部括弧而且合并同类项，我们得到

$$A'x''^2+C'y''^2+2(A'\xi'+D')x''$$
$$+2(C'\eta'+E')y''+F'=0, \tag{10'}$$

这里我们用 F' 表示所有常数项的和；对于它是怎样一个数，我们现在并不感兴趣．

我们来讨论三种可能情形．

I. A' 和 C' 都不等于零．在这情形里，取 $\xi'=-\frac{D'}{A'}$，$\eta'=-\frac{E'}{C'}$，我们就消去了一次项 x'' 和 y''，而且得到下列形状的方程：

$$A'x''^2+C'y''^2+F'=0. \tag{I}$$

II. $A'\neq 0$，$C'=0$，但是 $E'\neq 0$．取 $\xi'=-\frac{D'}{A'}$，$\eta'=0$，即

$y''=y'$, 我们得到方程

$$A'x''^2+2E'y'+F'=0,$$

或者

$$A'x''^2+2E'\left(y'+\frac{F'}{2E'}\right)=0.$$

再沿 Oy' 轴平行移动一个量 $\eta''=-\frac{F'}{2E'}$，我们求得 $y'=y''-\frac{F'}{2E'}$，即 $y'+\frac{F'}{2E'}=y''$，于是我们得到方程

$$A'x''^2+2E'y''=0. \tag{II}$$

如果 $A'=0$, $C'\neq 0$ 和 $D'\neq 0$, 我们只要交换 x 和 y 的地位，就得到上述情形.

III. $A'\neq 0$, $C'=0$, $E'=0$. 仍然取 $\xi'=-\frac{D'}{A'}$, $\eta'=0$, 我们得到方程

$$A'x''^2+F'=0. \tag{III}$$

如果 $A'=0$, $C'\neq 0$, $D'=0$, 我们仍然可以交换 x 和 y 的地位.

以上已经穷尽了所有的可能性，因为 A' 和 C' 不能同时等于零，否则就会降低方程的次数，而我们知道在坐标变换下这次数是不会改变的.

总之，经过直角坐标的适当选取，每一个二次方程可以化成三种所谓"归范"方程(I)，(II)和(III)中的一个.

设方程有形状(I)(这种情形里 A' 和 C' 都不等于零). 如果 $F'\neq 0$, 则可把方程(I)写成

$$\frac{x''^2}{-\frac{F'}{A'}}+\frac{y''^2}{-\frac{F'}{C'}}-1=0,$$

然后，根据 A', C', F' 的符号，把它化成方程(1)，(2)和(4)中的一个. 如果 x''^2 的分母是负数，而 y''^2 的分母是正数，则还需要改变坐标轴的名称，改 $O''x''$ 为 $O''y''$ 和改 $O''y''$ 为 $O''x''$.

如果 $F'=0$, 则在把方程(I)改写成

$$\frac{x''^2}{\frac{1}{A'}}+\frac{y''^2}{\frac{1}{C'}}=0$$

以后，我们引出方程(3)或者(5)。

如果方程有形状(II)(在这种情形里 A' 和 E' 都不等于零)，则在把它写成

$$x''^2+\frac{2E'}{A'}\cdot y''=0$$

以后，只要把 $-\frac{E'}{A'}$ 记做 p 而且将轴 $O''x''$ 和 $O''y''$ 的名称改为 $O''y''$ 和 $O''x''$，我们就得到方程(6)。

最后，如果我们有形状(III)的方程(并且这时 $A'\neq0$)，则它可以改写成

$$x''^2+\frac{F'}{A'}=0,$$

我们就得到方程(7)，(8)和(9)中的一个。

刚才证明了的，关于每一个二次方程都可以化成9种标准形状中的一种的定理，欧拉就已经详细地探讨过了。欧拉书中的论证只是在形式上与刚才所引的有些不同罢了。

§9. 用三个数规定力、速度和加速度。向量理论

紧接着欧拉以后，拉格朗日作了重要的一步。1788年拉格朗日在他著的《解析力学》里把力、速度和加速度算术化了，完全像笛卡儿把点算术化一样。拉格朗日引入他的书中的这个观念，后来以所谓向量理论的形式出现，是在物理、力学和技术中的极重要的助手。

空间中的直角坐标 首先我们要指出，不论是笛卡儿还是牛顿都没有能在空间中展开解析几何。后来在十八世纪前半叶，这就已经由克雷洛和拉盖尔做到了。为了在空间中规定一个点 M，我们取三条互相垂直的轴 Ox, Oy 和 Oz，而且考虑点 M 到平面 Oyz, Oxz 和 Oxy 的距离，这些距离取对应的符号就叫做点 M 的横坐标 x、纵坐标 y 和立坐标 z。

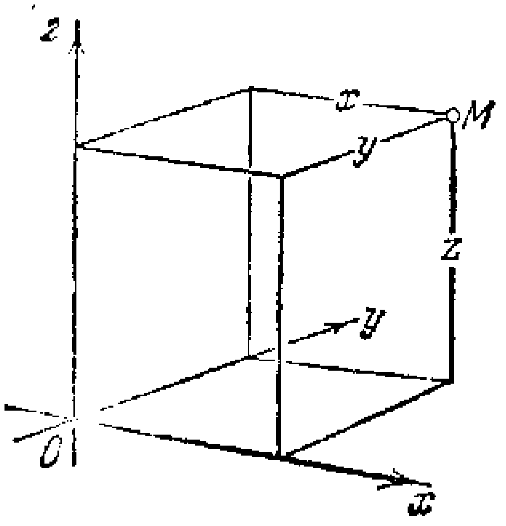


图 39

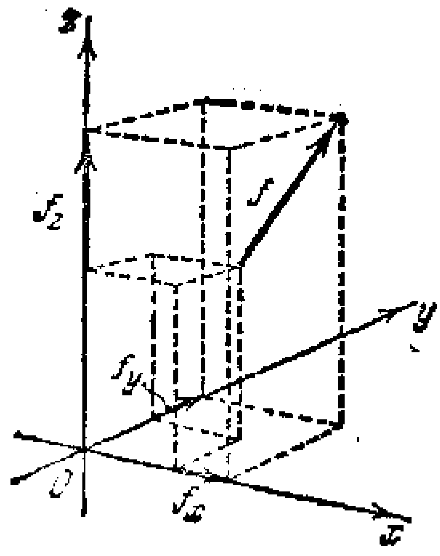


图 40

拉格朗日引入的力、速度和加速度的算术化 我们来讨论一个力 f(图 40). 在取定单位后，它可以用具有确定的长度和方向的有向线段来表示. 拉格朗日说，这个力 f 可以分解成沿着轴 Ox, Oy 和 Oz 的三个分力：f_x, f_y 和 f_z; 这些分力作为轴上的有向线段，已经可以简单地用数来给定它们，这些数的正负依赖于有向线段的方向是否与轴的正方向相同. 因此就可以讨论例如力 (2, 3, 4) 或者力 (1, −2, 5) 等等. 当力按平行四边形定律相加时，容易证明(后面将会做到这一点)，它们的对应的分力也相加. 举例说，上面写过的两力之和是力 (2+1, 3−2, 4+5) = (3, 1, 9). 对于速度和加速度也可以做同样的工作，在力学的全部问题中，可以把联系力、速度和加速度的所有力学方程，写成联系它们的分力的方程，即单是联系数的方程；只是每个方程都得写成三个方程：一个关于 x 的，第二个关于 y 的，第三个关于 z 的.

正好在拉格朗日以后一百年，数学和物理在当时发展着的电学理论的影响之下，开始广泛地讨论这种有确定长度和方向的线段的一般理论. 这种线段就被称为向量.

向量理论在力学、物理和技术方面有很大的价值，而它的叫做向量代数(以别于向量分析)的代数部分，现在已经是解析几何的重要组成部分.

向量代数 任意的有向线段(不论它表示的是:力、速度、加速度还是任何别的东西),即有已知长度和确定方向的线段,叫做**向量**.有同一长度和同一方向的两个向量叫做**相等的**,即在"向量"概念中考虑的只是它的长度和它的方向.向量可以相加.设给了向量 $\boldsymbol{a}, \boldsymbol{b}, \cdots, \boldsymbol{d}$. 从一个点引出向量 $\boldsymbol{a}$, 然后从向量 $\boldsymbol{a}$ 的终点引出向量 $\boldsymbol{b}$, 等等.我们得到了所谓向量折线 $\boldsymbol{a}\boldsymbol{b}\cdots\boldsymbol{d}$(图 41).向量 $\boldsymbol{m}$, 其起点与这折线的第一个向量 $\boldsymbol{a}$ 的起点重合,而其终点则与这折线的最后一个向量 $\boldsymbol{d}$ 的终点重合,这向量就叫做这些向量的**和**.

$$\boldsymbol{m}=\boldsymbol{a}+\boldsymbol{b}+\cdots+\boldsymbol{d}. \tag{11}$$

容易证明,向量 $\boldsymbol{m}$ 与各项 $\boldsymbol{a}, \boldsymbol{b}, \cdots, \boldsymbol{d}$ 选取的次序无关.

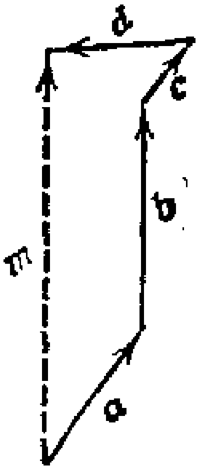


图 41

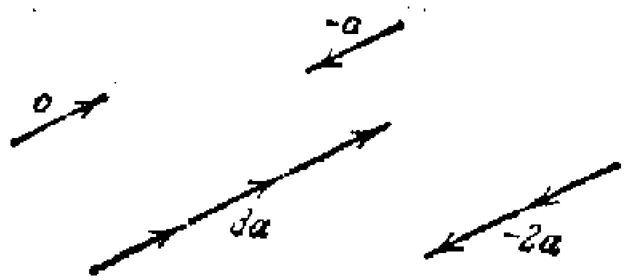


图 42

长度与向量 $\boldsymbol{a}$ 的长度相等,而方向则与它相反的向量,叫做它的**反向量**,而且记做 $-\boldsymbol{a}$.

减去向量 $\boldsymbol{a}$ 就是指加上它的反向量.

普通的实数在向量计算里叫做数量,设给了向量 $\boldsymbol{a}$(图 42)和数量 λ, 那末向量 $\boldsymbol{a}$ 乘上数量(数)λ 的乘积,即 $\lambda\boldsymbol{a}$, 是指这样一个向量,其长度等于向量 $\boldsymbol{a}$ 的长度 $|\boldsymbol{a}|$ 乘上数 λ 的绝对值,而其方向当 $\lambda>0$ 时与 $\boldsymbol{a}$ 相同,当 $\lambda<0$ 时与 $\boldsymbol{a}$ 相反.

我们来讨论笛卡儿直角坐标系 $Oxyz$ 和向量 $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$, 这些向量都有单位长度而且方向分别与轴 Ox, Oy, Oz 的正方向相同.明显地,从原点 O 到空间中任意取的点 M(图 43),可以先沿着向

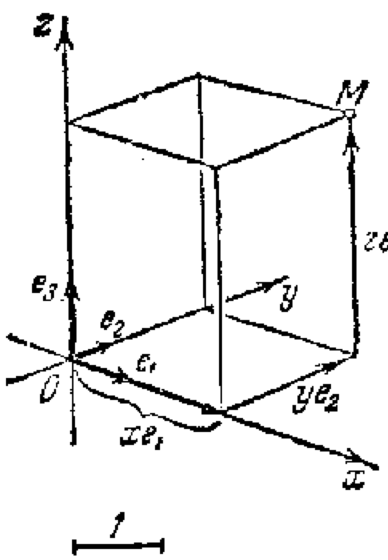


图 43

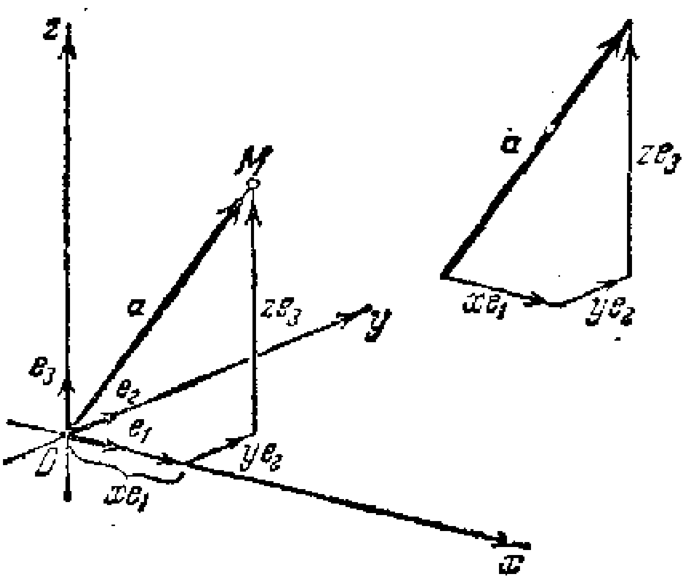


图 44

量 $\boldsymbol{e}_1$ 走若干"步"(正负整数、分数或无理数"步数"), 然后沿着向量 $\boldsymbol{e}_2$ 走若干"步", 最后沿着向量 $\boldsymbol{e}_3$ 走若干"步". 表明在这时沿向量 $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ 必须走若干"步"的数 x, y, z, 显然就是点 M 的笛卡儿坐标.

设给了一个向量 $\boldsymbol{a}$; 我们让一个点从这向量的起点移动到它的终点, 而且把这个移动分解成平行于轴 Ox, Oy 和 Oz 的移动, 于是如果这时这个点平行于 Ox 轴移动 $x\boldsymbol{e}_1$, 平行于 Oy 轴移动 $y\boldsymbol{e}_2$, 平行于 Oz 轴移动 $z\boldsymbol{e}_3$, 则

$$\boldsymbol{a}=x\boldsymbol{e}_1+y\boldsymbol{e}_2+z\boldsymbol{e}_3. \tag{12}$$

数 x, y, z 叫做向量 $\boldsymbol{a}$ 的坐标. 当把向量 $\boldsymbol{a}$ 的起点放在坐标原点 O 上时, 这显然就是这个向量的终点 M 的坐标(图 44). 由此可知, 当向量相加时它们的同名坐标相加, 相减时它们的同名坐标相减. 如果第一个向量沿 Ox 轴"挪动" $x\boldsymbol{e}_1$, 第二个向量沿 Ox 轴"挪动" $x'\boldsymbol{e}_1$, 则它们之和显然沿 Ox 轴"挪动" $(x+x')\boldsymbol{e}_1$, 等等(图45). 还可以知道, 当向量乘上某个数时, 它的坐标也乘上这个数.

数量乘积和它的性质 如果给了两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$, 则等于它们的长度乘上它们之间角 φ 的余弦的乘积 $|\boldsymbol{a}||\boldsymbol{b}|\cos\varphi$ 的数, 就叫做它们的数量乘积, 而且记做 $\boldsymbol{ab}$ 或者$(\boldsymbol{ab})$. 设(x, y, z)是向量 $\boldsymbol{a}$ 的坐标, $(\bar{x}, \bar{y}, \bar{z})$是向量 $\boldsymbol{b}$ 的坐标, 这时候数量乘积就等于

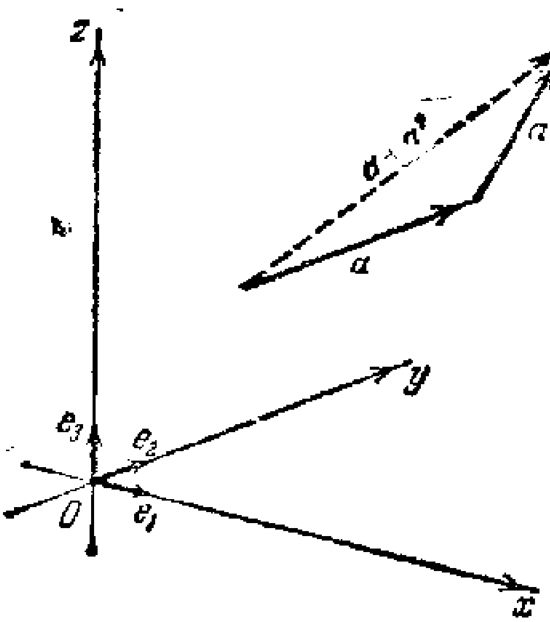


图 45

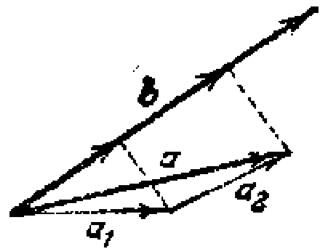


图 46

$$\boldsymbol{ab}=x\bar{x}+y\bar{y}+z\bar{z}, \tag{13}$$

即等于它们的同名坐标的乘积之和.

这个重要公式可以这样地来证明. 我们先作以下的注解:

1° 如果我们让数量乘积中的一个向量(例如向量 $\boldsymbol{a}$) 乘上一个数 λ, 则整个数量乘积显然也乘上这同一个数, 即

$$(\lambda\boldsymbol{a})\boldsymbol{b}=\lambda(\boldsymbol{ab}).$$

2° 数量乘积是可分配的, 即如果向量 $\boldsymbol{a}=\boldsymbol{a}_1+\boldsymbol{a}_2$, 则 $\boldsymbol{ab}=\boldsymbol{a}_1\boldsymbol{b}+\boldsymbol{a}_2\boldsymbol{b}$.

实际上, 这个等式的左边等于向量 $\boldsymbol{b}$ 的长度乘上向量 $\boldsymbol{a}$ 在向量 $\boldsymbol{b}$ 的轴上的投影数值(图 46), 而右边则等于向量 $\boldsymbol{b}$ 的长度乘上向量 $\boldsymbol{a}_1$ 和 $\boldsymbol{a}_2$ 在向量 $\boldsymbol{b}$ 的轴上的投影数值之和. 但是[1]

$$\text{пр. }\boldsymbol{a}=\text{пр. }\boldsymbol{a}_1+\text{пр. }\boldsymbol{a}_2,$$

这就证明了等式的正确性.

现在设给了两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$, 而且它们对向量 $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ 的分解式是 $\boldsymbol{a}=x\boldsymbol{e}_1+y\boldsymbol{e}_2+z\boldsymbol{e}_3$, $\boldsymbol{b}=\bar{x}\boldsymbol{e}_1+\bar{y}\boldsymbol{e}_2+\bar{z}\boldsymbol{e}_3$, 那末

$$\boldsymbol{ab}=(x\boldsymbol{e}_1+y\boldsymbol{e}_2+z\boldsymbol{e}_3)(\bar{x}\boldsymbol{e}_1+\bar{y}\boldsymbol{e}_2+\bar{z}\boldsymbol{e}_3).$$

根据分配性(2°), 括弧里的向量之和的数量乘积可以像多项式那

1) пр 是俄文 проекция (投影)的前两个字母, пр. $\boldsymbol{a}$ 代表向量 $\boldsymbol{a}$ 的投影数值. ——译者注

样相乘，而根据(1°)，在所得到的每一项里都可以把数量因子放在前面，所以

$$\begin{aligned}\boldsymbol{a}\boldsymbol{b}=&x\bar{x}\boldsymbol{e}_1\boldsymbol{e}_1+x\bar{y}\boldsymbol{e}_1\boldsymbol{e}_2+x\bar{z}\boldsymbol{e}_1\boldsymbol{e}_3+y\bar{x}\boldsymbol{e}_2\boldsymbol{e}_1+y\bar{y}\boldsymbol{e}_2\boldsymbol{e}_2\\&+y\bar{z}\boldsymbol{e}_2\boldsymbol{e}_3+z\bar{x}\boldsymbol{e}_3\boldsymbol{e}_1+z\bar{y}\boldsymbol{e}_3\boldsymbol{e}_2+z\bar{z}\boldsymbol{e}_3\boldsymbol{e}_3.\end{aligned}$$

但是 $|\boldsymbol{e}_1|=|\boldsymbol{e}_2|=|\boldsymbol{e}_3|=1,\quad \cos 0°=1$ 和 $\cos 90°=0.$

因此

$$\begin{aligned}&\boldsymbol{e}_1\boldsymbol{e}_1=1,\quad \boldsymbol{e}_1\boldsymbol{e}_2=0,\quad \boldsymbol{e}_1\boldsymbol{e}_3=0,\\&\boldsymbol{e}_2\boldsymbol{e}_1=0,\quad \boldsymbol{e}_2\boldsymbol{e}_2=1,\quad \boldsymbol{e}_2\boldsymbol{e}_3=0,\\&\boldsymbol{e}_3\boldsymbol{e}_1=0,\quad \boldsymbol{e}_3\boldsymbol{e}_2=0,\quad \boldsymbol{e}_3\boldsymbol{e}_3=1.\end{aligned}$$

这样一来，

$$\boldsymbol{a}\boldsymbol{b}=x\bar{x}+y\bar{y}+z\bar{z}. \tag{14}$$

我们注意到，如果向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 互相垂直，则 $\varphi=90°$，$\cos\varphi=0$，因而有等式

$$x\bar{x}+y\bar{y}+z\bar{z}=0. \tag{15}$$

这就是容易验证的向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 互相垂直的条件.

两个方向之间的角 我们来讨论由其与坐标轴的夹角 α, β, γ 决定的方向. 通过坐标原点 O 引这个方向的直线，而且在这直线上截取单位长度的线段 OA(图 47). 在这情形下，点 A 的坐标(即向量 $\overrightarrow{OA}$ 的坐标)正是 $\cos\alpha$, $\cos\beta$, $\cos\gamma$. 如果还有由角 $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$ 给定的第二个方向，则关于这第二个方向的类似的向量 $\overrightarrow{OB}$ 就有坐标 $\cos\bar{\alpha}$, $\cos\bar{\beta}$, $\cos\bar{\gamma}$(图 48). 设 φ 是这两个向量之间的

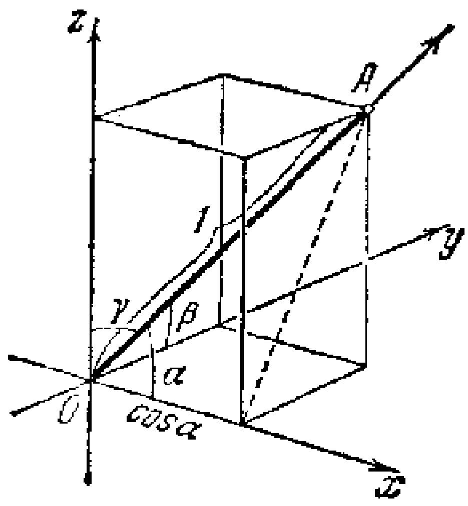


图 47

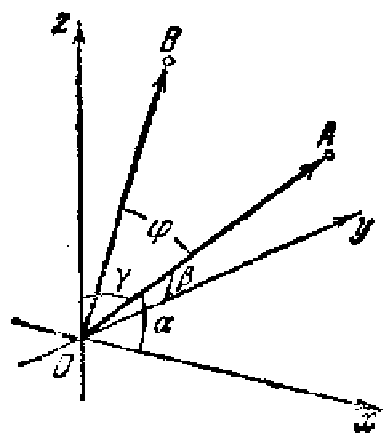


图 48

角，那末它们的数量乘积就等于 $1\cdot 1\cos\varphi$，于是我们求得

$$\cos\varphi=\cos\alpha\cos\bar{\alpha}+\cos\beta\cos\bar{\beta}+\cos\gamma\cos\bar{\gamma}. \tag{16}$$

这是关于两个方向之间的角的余弦的重要公式.

§ 10. 空间解析几何. 空间中的曲面的方程和曲线的方程

如果给了方程 $z=f(x, y)$，而且认为 x 和 y 分别是点的横坐标和纵坐标，z 是立坐标，则这个方程就表示一个曲面 P，从平面 Oxy 的点 (x, y) 上引出等于 z 的垂直线段，就可以得到这个曲面. 这些垂直线段的端点的几何轨迹就给出由这方程表示的曲面. 如果联系 x, y 和 z 的方程并未对 z 解出，则可以先把 z 解出，然后再来作出这个曲面 P. 一般在解析几何里，所谓由带三个变数 x, y, z 的方程表示的曲面，指的是空间中坐标 x, y, z 满足这个方程的全体点的集合(图 49).

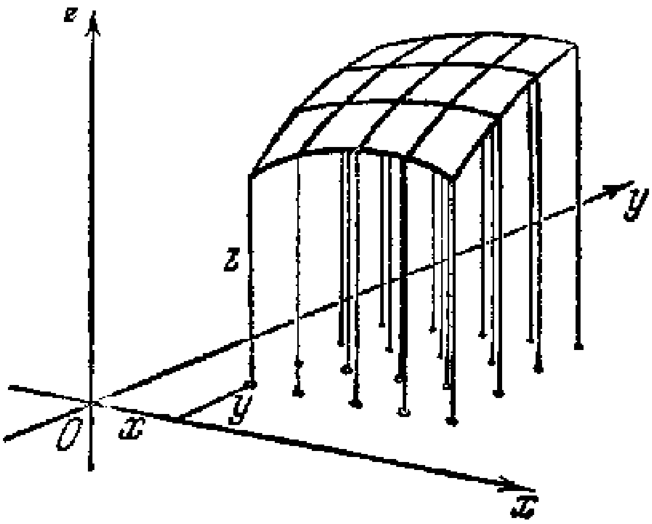


图 49

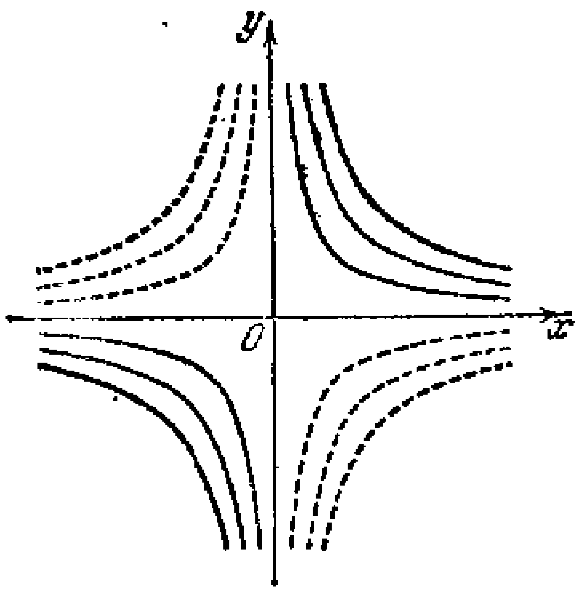


图 50

在第二章里就曾经说过，两个变数的函数 $f(x, y)$，不仅可以描述曲面 P，而且可以描述这曲面的水平线组，那就是让函数 $f(x, y)$ 取常数值时在平面 Oxy 上所表示的曲线. 这曲线组显然

不是别的，正是曲面 P 在平面 Oxy 上的地形图.

例 方程 $xy=z$ 给出例如这样的水平线：…，$xy=-3$，$xy=-2$，$xy=-1$，$xy=0$，$xy=1$，$xy=2$，$xy=3$，… 它们全都是双曲线(图 50)，除掉曲线 $xy=0$，这曲线正是一对坐标轴，得到的显然是一个马鞍形曲面(图 51)(所谓双曲抛物面).

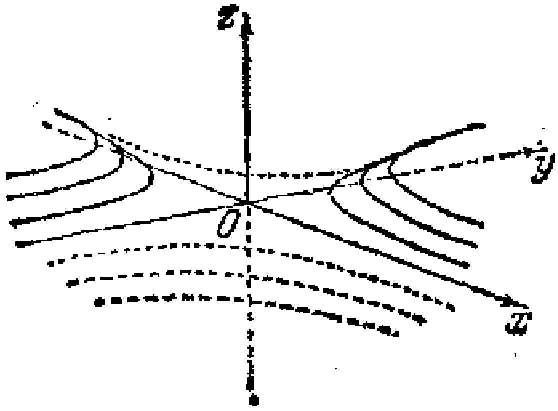


图 51

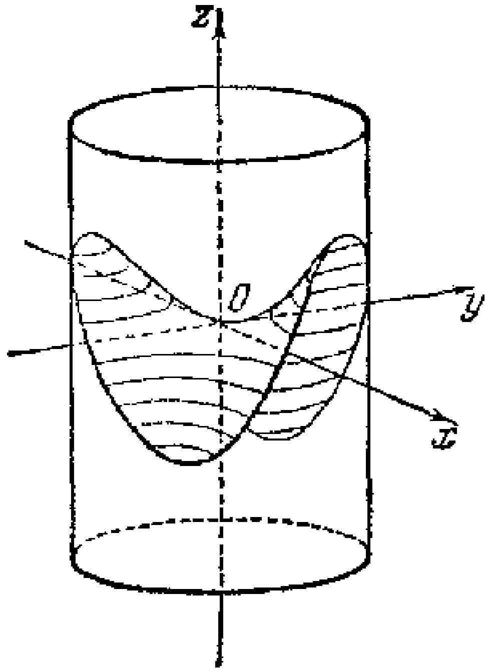


图 52

为了给定空间中的曲线，可以用相交于这曲线的任何两个曲面 P 和 Q 的方程. 譬如说，方程组

$$xy=z,$$
$$x^2+y^2=1$$

给定一条空间曲线(图 52). 方程 $xy=z$ 决定刚才讨论过的双曲抛物面，方程 $x^2+y^2=1$ 决定一个圆柱面，它有单位长的半径，它的轴就是 Oz 轴. 因此所说的方程组给定抛物面与圆柱面的交线，就画在图 52 上.

如果在这方程组里随意地取一个未知数(例如 x)的值，然后对 y 和 z 来解这个组，则得到的就是这曲线上各个点的坐标.

平面的方程和直线的方程 可以证明，每一个带三个变数的一次方程

$$Ax+By+Cz+D=0$$

表示一个平面，反过来也对．根据以上所说，显然直线可以用一组两个这种方程来给定：

$$A_1x+B_1y+C_1z+D_1=0,$$
$$A_2x+B_2y+C_2z+D_2=0,$$

即作为两个平面的交线而给定．

带三个变数的一般的二次方程和它的十七个标准形状 带三个变数的二次方程

$$A_1x^2+A_2y^2+A_3z^2+2B_1yz+2B_2xz+2B_3xy$$
$$+2C_1x+2C_2y+2C_3z+D=0 \tag{17}$$

包含10项．就像对于带两个变数的方程所曾经做过的一样，可以证明，对应地让给定的直角坐标系绕原点旋转以后，可以把方程(17)化成

$$A_1'x'^2+A_2'y'^2+A_3'z'^2+2C_1'x'+2C_2'y'+2C_3'z'+D=0, \tag{18}$$

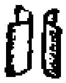
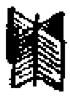
即消去了变数乘积的项．只是与平面的情形相比，证明这样地简化方程的可能性现在要困难得多．证明的难点就在于：平面上绕点的旋转由一个角 φ 决定，我们那时就选取了一个角．然而在空间中，立体绕不动点的旋转由三个不相关的角(欧拉角) φ, θ, ψ 决定，并且是十分复杂的．所以必须通过迂回的途径(参看第三卷第十六章里利用正交变换把二次齐式化成平方和的理论)来证明在方程中消去变数乘积项的可能性．然后，也象在平面上一样，再作某个平行移轴而且把方程约简，此后方程(18)终于具有下列标准形状中的一个：

1) $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}+\dfrac{z^2}{c^2}-1=0$  椭圆面

2) $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}+\dfrac{z^2}{c^2}+1=0$  虚椭圆面

3) $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}-\dfrac{z^2}{c^2}-1=0$  单叶双曲面

4) $\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}-\dfrac{z^2}{c^2}+1=0$  双叶双曲面

5) $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0$		二次锥面
6) $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=0$		虚二次锥面
7) $\frac{x^2}{a^2}+\frac{y^2}{b^2}-2cz=0$		椭圆抛物面
8) $\frac{x^2}{a^2}-\frac{y^2}{b^2}-2cz=0$		双曲抛物面
9) $\frac{x^2}{a^2}+\frac{y^2}{b^2}-1=0$		椭圆柱面
10) $\frac{x^2}{a^2}+\frac{y^2}{b^2}+1=0$		虚椭圆柱面
11) $\frac{x^2}{a^2}+\frac{y^2}{b^2}=0$		一对虚的相交平面
12) $\frac{x^2}{a^2}-\frac{y^2}{b^2}-1=0$		双曲柱面
13) $\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$		一对相交的平面
14) $y^2-2px=0$		抛物柱面
15) $x^2-a^2=0$		一对平行的平面
16) $x^2+a^2=0$		一对虚的平行平面
17) $x^2=0$		一对重合的平面

最后九个标准方程(9)—(17)不包含带 z 的项而且正是平面 Oxy 上二次曲线的标准方程. 在空间中这些方程都表示柱面，它们的导线是在平面 Oxy 上的对应的二次曲线，它们的母线平行于 Oz 轴. 实际上，如果这些方程中的任何一个被坐标$(x_1, y_1, 0)$的点所满足，则它也被坐标(x_1, y_1, z)的点所满足，这时 z 可以任意地取值，因为在方程中根本没有带 z 的项.

在方程(1)—(8)中容易看出，方程(2)不被任何有实数坐标(x, y, z)的点所满足，而方程(6)则只被一个这样的点(0, 0, 0)——坐标原点所满足. 因此，剩下只要研究六个方程：(1)，(3)，(4)，

(5), (7), (8).

椭圆面 我们来比较由方程 $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}-1=0$ 表示的曲面和由方程 $x^2+y^2+z^2-1=0$ 表示的曲面. 其中第二个方程显然是一个球面 C 的方程, 这球面以坐标原点为中心而且半径等于1, 这是因为 $x^2+y^2+z^2$ 是从点 (x, y, z) 到原点 O 的距离的平方. 如果 (x, y, z) 是球面 C 上的点, 即满足第二个方程, 则 (ax, by, cz) 是坐标满足第一个方程的点. 因此, 由第一个方程表示的曲面可以这样地从球面 C 得到, 只要把球面上点的所有横坐标 x 换成 ax, 纵坐标 y 换成 by, 立坐标 z 换成 cz, 即只要让球面 C 作背着平面 Oyz, Oxz 和 Oxy 而且有伸展系数 a, b 和 c 的均匀伸展. 这个曲面叫做椭圆面(图 53).

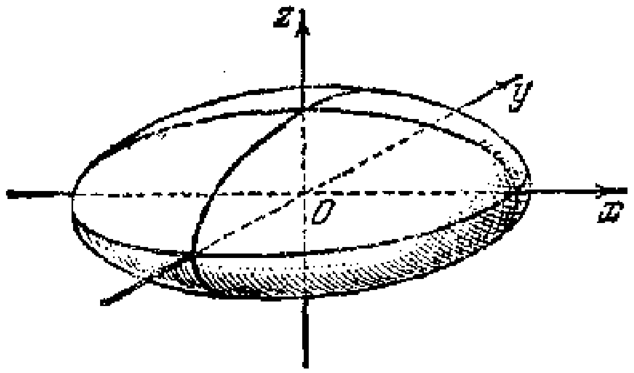


图 53

双曲面和二次锥面 现在我们来讨论方程(3), (4)和(5), 即讨论下列形状的方程:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=\delta, \tag{19}$$

这里 $\delta=1$, -1 或 0. 我们来比较这个方程与方程

$$\frac{x^2}{a^2}+\frac{y^2}{a^2}-\frac{z^2}{c^2}=\delta, \tag{20}$$

这后一个方程中 y^2 的分母也是 a^2, 而不像方程(19)中是 b^2. 与以前作过的注解类似, 曲面(19)从曲面(20)经过背着平面 Oxz 作系数 b/a 的伸展而得到.

现在让我们来看一下曲面(20)是什么. 我们取垂直于 Oz 轴的任何平面 $z=h$, 来研究它与曲面(20)的交线. 把 $z=h$ 代入方程(20), 我们得到方程

$$x^2+y^2=a^2\left(\delta+\frac{h^2}{c^2}\right).$$

如果 $\delta+\frac{h^2}{c^2}$ 是正数，则得到的方程与方程 $z=h$ 共同给出一个圆，这个圆处在平面 $z=h$ 上而且有中心在 Oz 轴上. 如果 $\delta+\frac{h^2}{c^2}$ 是负数(这只有当 $\delta=-1$ 而且 h^2 较小时才可能)，则平面 $z=h$ 根本不与曲面(20)相交，因为平方和 x^2+y^2 不能是负数.

因此，整个曲面(20)由圆组成，这些圆处在垂直于 Oz 轴的平面上而且有中心在 Oz 轴上. 但是在这种情形下，曲面(20)是绕 Oz 轴的旋转曲面. 于是只要让它与通过 Oz 轴的任何平面相交来求出它的"经线"，即求出处在通过旋转轴的平面上的、由它来转出曲面的曲线.

我们让曲面(20)与坐标平面 Oxz(即平面 $y=0$)相交(图 54)，把 $y=0$ 代入方程(20)，我们就得到经线的方程

$$\frac{x^2}{a^2}-\frac{z^2}{c^2}=\delta.$$

当 $\delta=1$ 时它是双曲线 I，当 $\delta=-1$ 时它是双曲线 II，当 $\delta=0$ 时它是一对相交的直线 III. 在旋转下它们给出所谓单叶旋转双曲面(图 55)、双叶旋转双曲面(图 56)和正圆锥面(图 57).

抛物面 还剩下方程(7)和(8). 我们来比较其中的第一个方程

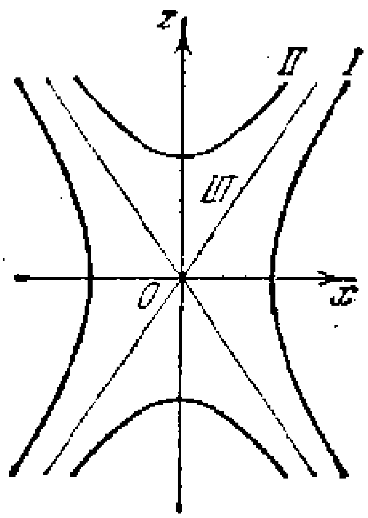


图 54

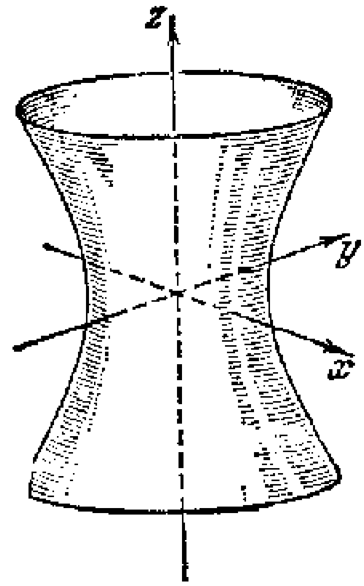


图 55

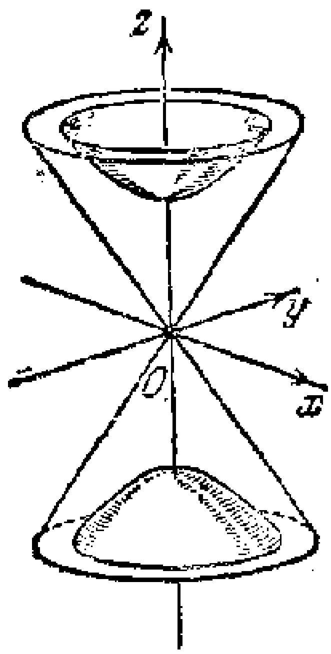


图 56

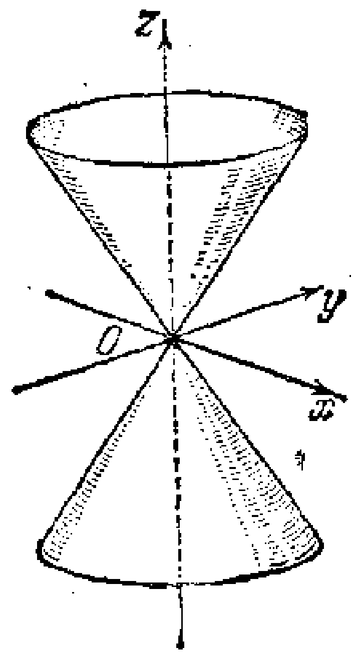


图 57

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=2cz$$

与方程

$$\frac{x^2}{a^2}+\frac{y^2}{a^2}=2cz,$$

后者经过与上段相类似的研究可以肯定，它表示的是由抛物线 $x^2=2a^2cz$ 绕 Oz 轴旋转而得到的曲面——所谓旋转抛物面(图 58)，我们在谈论抛物面镜时已经提到它了．一般的椭圆抛物面(7)从旋转抛物面经过背着平面 Oxz 的伸展而得到．

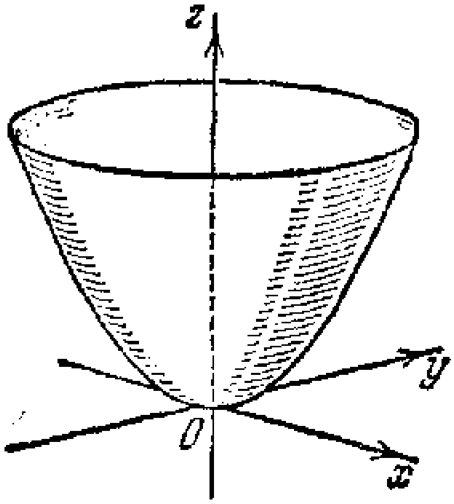


图 58

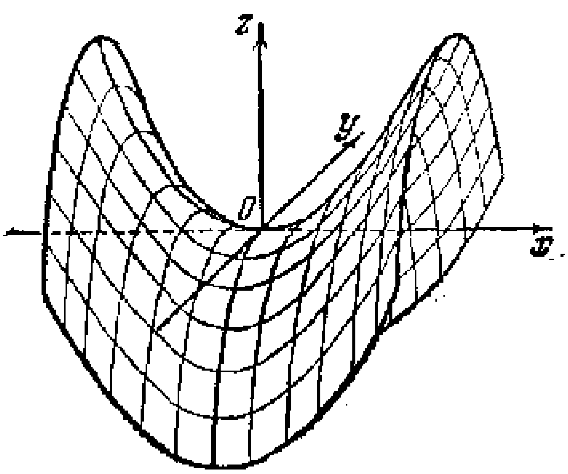


图 59

曲面(8)必须用另一种方式来研究：例如研究它与平面 $z=h$ 的交线，它们都是双曲线。曲面(8)的水平线图画在图 50 上；我们还在图 51 上看到这曲面对于坐标轴的另一个位置。它有图 59 上所画的形状，而且叫做双曲抛物面。它与平行于平面 Oxz 的平面的交线是相同的抛物线。平行于平面 Oyz 的平面也截出相同的截线。

单叶双曲面的直母线 单叶双曲面和双曲抛物面也像锥面和柱面一样，可以从移动直线而得到，这是非常奇妙但是并不十分显然的。在双曲面的情形，这只要对于单叶旋转双曲面 $\frac{x^2}{a^2}+\frac{y^2}{a^2}-\frac{z^2}{c^2}=1$ 证明就成了，因为一般的单叶双曲面从它作背着平面 Oxz 的均匀伸展而得到，而在这种伸展下，任意直线都还变成直线。让旋转双曲面与平行于平面 Oxz 的平面 $y=a$ 相交。用 $y=a$ 代入，我们得到

$$\frac{x^2}{a^2}+\frac{a^2}{a^2}-\frac{z^2}{c^2}=1 \quad 或者 \quad \frac{x^2}{a^2}-\frac{z^2}{c^2}=0.$$

但是这个方程与方程 $y=a$ 共同给出平面 $y=a$ 上的一对直线：$\frac{x}{a}-\frac{z}{c}=0$ 和 $\frac{x}{a}+\frac{z}{c}=0$。

总之，我们已经发现，在双曲面上有着一对相交的直线。现在如果让双曲面绕 Oz 轴旋转，则这两条直线中的每一条直线显然都扫过整个双曲面(图 60)。

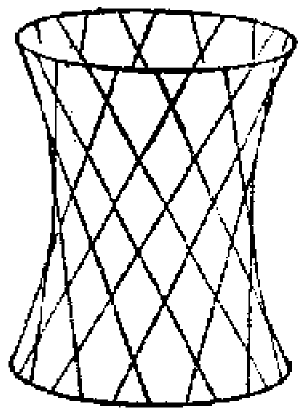


图 60

容易证明：1)得到的同一族中的任意两条直线都不在一个平面上(即所谓相错)，2)一族中的任意直线与另一族中的所有直线都相交(除掉与它相对的一条直线，那是与它平行的)，3)同一族中的任意三条直线都不平行于同一个平面。

利用火柴和针，容易得到关于单叶旋转双曲面的形象。如果用针穿过一根火柴的中部，而且把另一根火柴的中部穿在针的尖

端，使它与第一根火柴有平行的位置，则在以第一根火柴为轴作旋转时，第二根火柴画出一个圆柱面(图 61)．而如果第二根火柴垂直地穿在针上，但是不平行于第一根火柴，则它在同样的旋转下就画出一个旋转单叶双曲面，在转得很快时我们可以很好地看到它(图 62)．

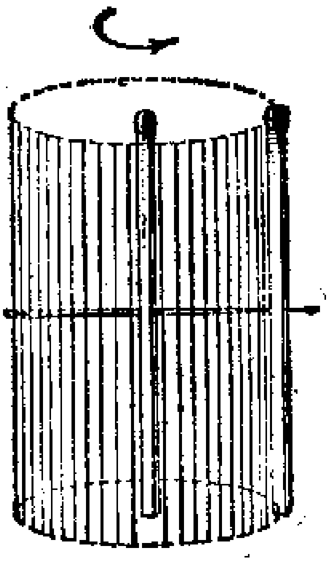


图 61

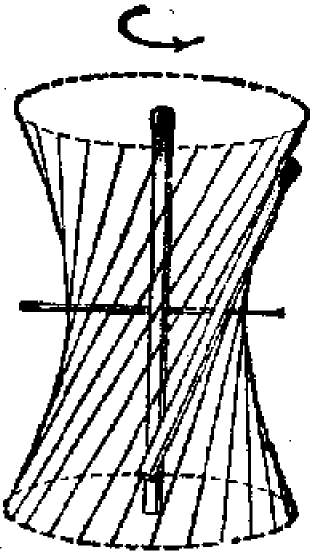


图 62

二次方程研究的总结　虽然一般的带三个变数的二次方程可以表示十七种本质上不同的曲面，但是要记住这些曲面并不困难．其中最后九种正是立在九种可能的二次曲线上的柱面．而前面八种则分成四对：两个椭圆面(实的和虚的)，两个双曲面(单叶的和双叶的)，两个二次锥面(实的和虚的)和两个抛物面(椭圆的和双曲的)．所有这些曲面在力学、物理和技术中都起着重要的作用．(惯性椭圆面，弹性椭圆面，物理学中罗仑兹变换里的双曲面，抛物面镜的旋转抛物面，等等．)

§11．仿射变换和正交变换

解析几何下一步重要的发展是在其中(而且一般地在几何学中)引入变换理论．这需要详细说明一下为什么．

平面向着直线的“压缩”　我们来讨论一个非常简单的平面变

换——向着直线的具有系数 k 的"压缩". 设在平面上给了直线 a 和一个正数 k, 例如 $k=\frac{2}{3}$. 我们让直线 a 的全部点留在原地, 而把不在这直线上的每一个点 M, 换成这样一个点 M', 点 M' 与点 M 处在直线 a 的同侧, 而且与 M 同在 a 的一条垂直线上, 但是从点 M' 到直线 a 的距离等于从点 M 到直线 a 的距离的 $\frac{2}{3}$. 如果系数 k 像刚才那样小于 1, 则作的是真正的平面向着直线的压缩; 而如果 k 大于 1, 则作的将是平面背着直线的伸展, 但是为了方便起见, 在这两种情况下我们都说是"压缩", 只是把"压缩"这个词放在引号里.

被变换的点和图形叫做原像, 而它所变换成的东西则叫做它的像. 举例说, 点 M' 是点 M 的像(图 63).

让我们来证明, 在平面向着直线的均匀"压缩"下, 平面上的每一条直线都变成直线. 实际上, 设平面以"压缩"系数 k 被"压"向处在这平面上的直线 a. 设 b 是平面上的任意直线, O 是它与直线 a 的交点, B 是它的任意别的点, BA 是从这个点引到直线 a 上的垂直线(图 64). 经过"压缩", 点 B 变成这垂直线上的一个点 B', 使 $B'A=k\cdot BA$. 所以角 $B'OA$ 的正切就等于 $\frac{AB'}{OA}=\frac{k\cdot AB}{OA}$, 即等于直线 b 与直线 a 的夹角的正切的 k 倍, 这就是说, 对于由直线 b 上不同的点变成的所有点 B' 说, 角 $B'OA$ 的正切都相同. 因此, 所有的点 B' 处在同一条直线上, 这直线通过点 O 而且与直线 a 的夹角有这样的正切.

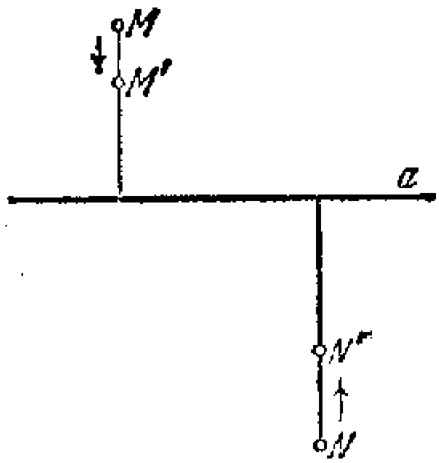


图 63

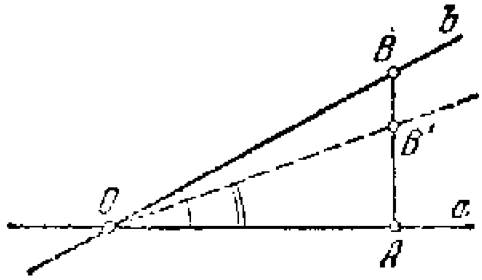


图 64

在"压缩"下，平行直线还成为平行直线．实际上，如果直线 b 和 c 与直线 a 的夹角有相同的正切，则它们的像 b' 和 c' 与 a 的夹角的正切因为都与上述正切相差一个因子 k，所以也彼此相同，即直线 b' 和 c' 也彼此平行．

平面上的每一个直线段在平面向着直线的"压缩"下均匀地缩短(或者伸长，虽然对于不同方向的线段它们的伸缩程度是不同的)．所谓"均匀"缩短，我们是说，线段的中点还是中点，三分点还是三分点，等等，这就是说，线段就其长度说均匀地受到压缩．实际上，不管点 M 分线段 M_1M_2 成什么比值，它的像 M' 分这线段的像 $M_1'M_2'$ 成同一个比值，这是因为平行直线(现在是直线 a 的垂直线)分它们的割线(这时是 b 和 b') 为成比例的部分(图 65)．

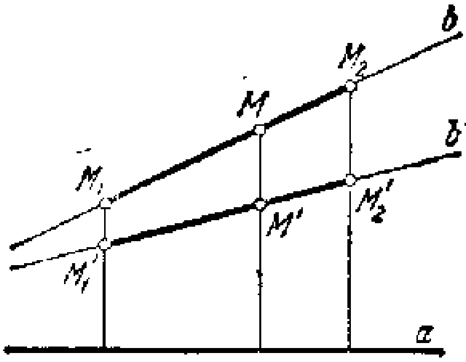


图 65

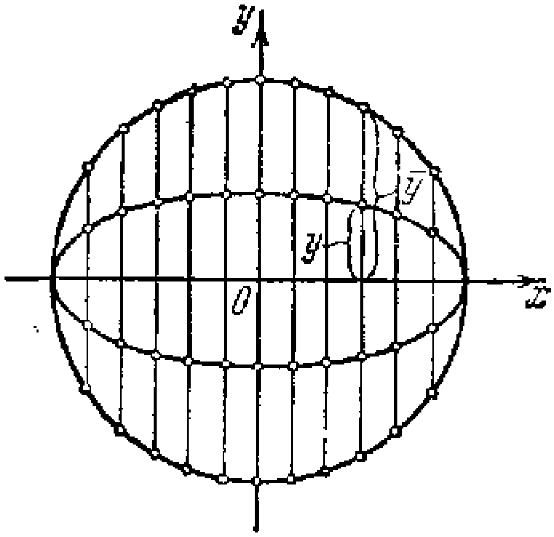


图 66

椭圆是圆经过"压缩"的结果　我们来讨论以原点为中心而且有半径 a 的圆．根据毕达哥拉斯定理，它的方程是 $x^2+\bar{y}^2=a^2$．我们不写 y，而写 $\bar{y}$，是因为我们在以后还要用到 y．我们来看一下，当平面向着 Ox 轴作系数 $\frac{b}{a}$ 的"压缩"时(图 66)，这个圆将会变成什么．经过这种"压缩"，所有点的 x 值保留原来的，而 $\bar{y}$ 值则变成等于 $y=\bar{y}\,\frac{b}{a}$，即 $\bar{y}=\frac{a}{b}\,y$．把 $\bar{y}$ 代入上面写出的圆的方程，我们就有：$x^2+\frac{a^2}{b^2}\,y^2=a^2$ 或者 $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$，这就是由所说的圆

经过向着 Ox 轴的压缩而得到的曲线在同样的坐标轴里的方程．我们看到，得出的是椭圆．总之，我们证明了，椭圆是圆经过"压缩"的结果．

由于椭圆是圆经过"压缩"的结果，由此直接得出椭圆的很多性质．举例说，前面提过直径的性质：如果给了椭圆的平行割线，则这些割线上的弦的中点处在一条直线上(图 12)．这个性质就可以用下列方法证明．我们作相反的伸展把椭圆变成圆．这时椭圆的平行弦变成圆的平行弦，而它们的中点则变成这些弦的中点．但是圆的平行弦的中点处在它的直径上，即处在一条直线上，因此，椭圆的平行弦的中点也在一条直线上．那就是说，当这个圆经过"压缩"变成椭圆时，圆的直径还变成一条直线，这些中点就处在这条直线上．

下面是"压缩"理论的另一个应用．因为圆的每一个铅垂的长条，当它压向 Ox 轴时，并不改变其宽度，而改变其长度为 b/a 倍，所以这长条的面积在压缩后等于原先的面积乘上 b/a，而因为圆的面积等于 πa^2，所以对应的椭圆的面积等于 $\pi a^2\dfrac{b}{a}=\pi ab$．

解更复杂的问题的例子 设给了一个椭圆，需要找出外切于这个椭圆的最小面积的三角形．我们先对圆来解这个问题．我们

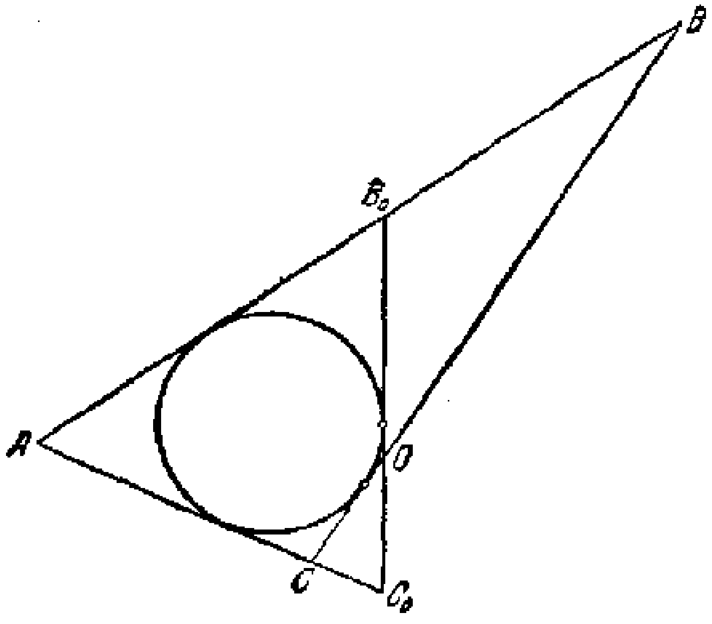


图 67

来证明，在圆的情形，这三角形是等边三角形．实际上，设外切三角形不是等边的，即它的最小角(我们把它记做 B)小于 $60°$，而最大角 $C>60°$．于是，不改变角 A，而把边 BC 变成边 B_0C_0 (图 67)，这时我们让顶点 B 向着 A 移动，直到角 B_0 和 C_0 中有一个变得等于 $60°$ 为止．我们得到一个有较小面积的外切三角形 AB_0C_0，因为这时[1]有 $OC<OB$，$OC_0\leqslant OB_0$，因而割下的面积 OBB_0 大于添上的面积 OCC_0．如果得到的三角形还不是等边的，则再一次重复全部的论证，我们还可以把它的面积再减小而化成等边三角形．因此，外切于已知圆的每一个不等边的三角形，都有比等边三角形大的面积．

现在我们转到椭圆．背着椭圆的长轴作这样的伸展，把它变成原来经过"压缩"而得到它的圆．在这种伸展下(图 68)：(1) 所有外切于椭圆的三角形，都变成外切于所得到的圆的三角形；(2) 所有图形的面积，特别是所有这些三角形的面积，都增大同一个倍数．由此我们看到，外切于椭圆的最小面积的三角形，就是那些由外切于圆的等边三角形变成的三角形，这种三角形有无穷多个，它们的重心都在椭圆的中心处，切点是它们各边的中点．从刚才提到过的圆出发，很容易作出每一个这种的三角形(图 68)．

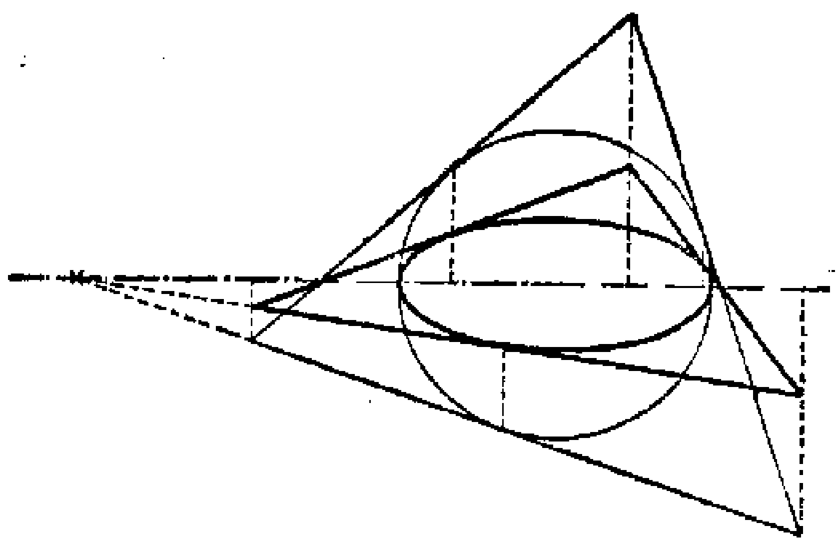


图 68

1) 这是易于证明的．

平面向着直线的"压缩"只不过是平面上更一般的所谓仿射变换的一种特殊情形.

一般的仿射变换 我们把从一个公共起点 O 引出的两个不在一条直线上的向量 e_1, e_2, 叫做平面上的坐标"架". 平面上的点 M 对于这种标架 Oe_1e_2 的坐标是指这样的数 x, y, 从原点 O 到点 M, 必须从点 O 放置向量 e_1x 次, 然后放置向量 e_2y 次. 这是平面上的一般的笛卡儿坐标. 同样可以定义空间中的一般的笛卡儿坐标. 我们在这以前所利用的普通所谓的直角笛卡儿坐标是它的一种特殊情形, 相当于坐标向量互相垂直, 而且它们的长度都等于我们取来测量所有线段的长度的尺度单位的长度.

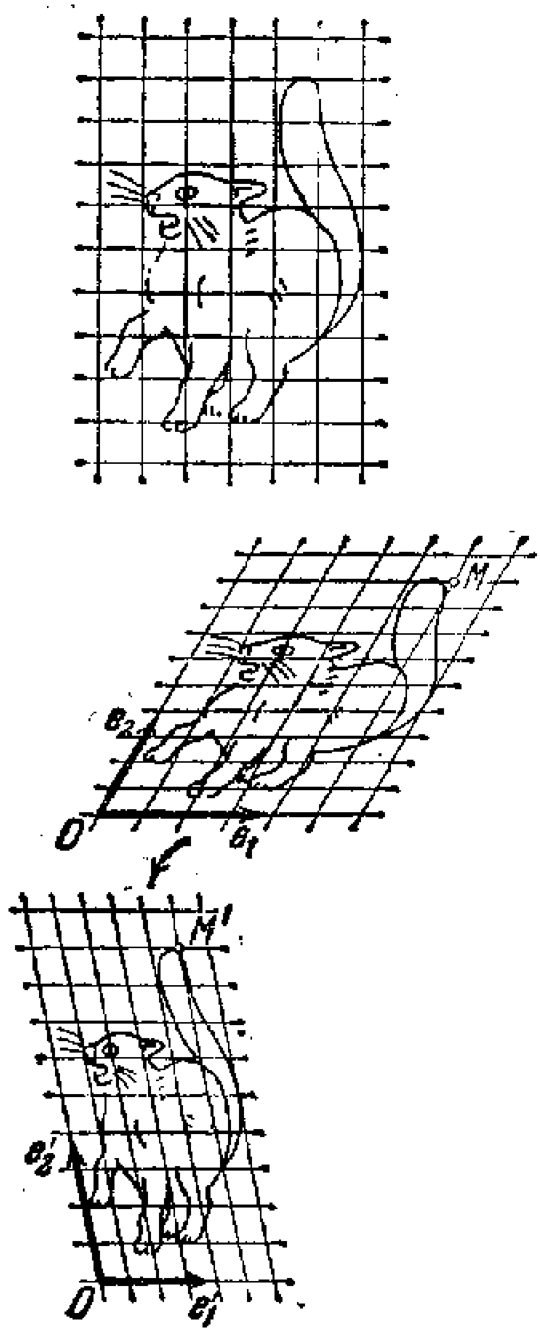


图 69

一般的平面仿射变换是这样的一种变换, 它把给定的相等平行四边形的网变成任何别的相等平行四边形的网. 确切地说, 它是这样的平面变换, 它把给定的坐标架 Oe_1e_2 变成一个别的标架(一般地说, 具有另外的"度量", 即其向量有另外的长度和另外的夹角), 而任意点 M 则变成这样的点 M', 点 M' 对于新标架的坐标与点 M 对于旧标架的坐标相同(图 69).

向着 Ox 轴的具有系数 k 的压缩是这样一种特殊的仿射变换,

它把直角标架 Oe_1e_2 变成标架 Oe_1ke_2.

很容易可以证明，在一般的仿射变换下也有：直线变成直线，平行直线变成平行直线，而且如果有一个点分一个线段成某个比值，则这点的像也分这线段的像成同一个比值．此外，可以证明一个值得注意的定理：可以这样地得出任何平面仿射变换，先让平面当做一个刚体沿自身作一次移动，然后向着两条互相垂直的直线作两个"压缩"，这两个"压缩"一般说具有不同的系数 k_1 和 k_2.

为了证明这个断言，我们来察看被变换的平面上一个圆的所有半径(图 70)．设半径 OA 是经过变换后变得最短的半径，设它变成 $O'A'$．于是 OA 的垂直线 AB 也变成 $O'A'$ 的垂直线 $A'B'$，因为如果另有与 $O'A'$ 不同的垂直线 $O'C'$，则它是倾斜线 OC 的像而且半径 OD 的像 $O'D'$ 是垂直线 $O'C'$ 的一部分，即短于倾斜线 $O'A'$，与假设相违背.

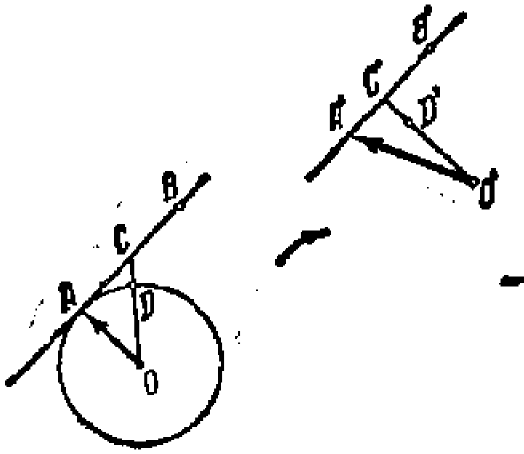


图 70

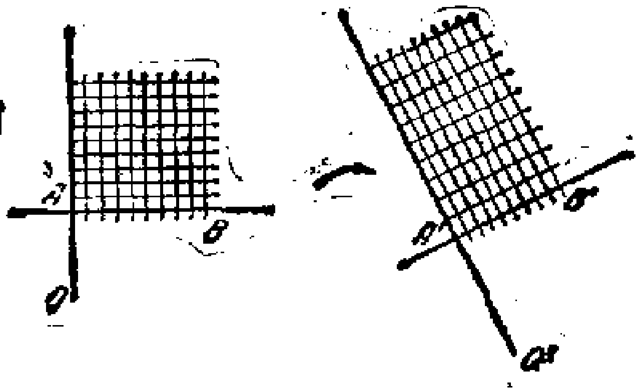


图 71

因而互相垂直的直线 OA 和 AB 变成互相垂直的直线 $O'A'$ 和 $A'B'$．因此，作在 OA 和 AB 上的正方形网变成一个相等的长方形网(图 71)，而且均匀"压缩"就沿着这个正方形网的直线而进行.

完全同样地可以这样来定义一般的空间仿射变换：它把空间的坐标架 $Oe_1e_2e_3$ 变成另外一个标架 $O'e_1'e_2'e_3'$，后者一般地说具有另外的"度量"，即坐标向量有另外的长度和另外的夹角；它还把任意点 M 变成一个点 M'，点 M' 对于新标架的坐标与点 M 对于

旧标架的坐标相同.

空间仿射变换也具有我们所列举过的全部性质，只是在最后的定理中，在空间情形提到的是空间作为刚体的移动，然后是向着三个互相垂直的平面的具有某些系数 k_1, k_2, k_3 的三个"压缩".

仿射变换的最重要的应用

1) 在几何学中用来解决有关于图形的仿射性质的问题. 所谓仿射性质是指在仿射变换下保留不变的性质. 关于椭圆直径的定理和关于外切三角形的问题就是这种问题的例子. 为了解决同类的问题，用仿射变换把一个图形变成一个比较简单的图形，在后一个图形中发掘所求的性质，然后再回到原来的图形.

2) 在解析几何学中用来作二次曲线和二次曲面的分类. 事实是：可以证明，不同的椭圆在经过仿射变换可以彼此变换的意义下是彼此同类的（仿射的拉丁原文 *affinis* 本来就是同类的意思）. 同样地，所有双曲线是彼此仿射的，所有抛物线也是彼此仿射的. 但是在任何仿射变换下，椭圆决不能变成双曲线或抛物线，双曲线也不能变成抛物线，这就是说，它们彼此是仿射的不同类的. 自然可以把所有二次曲线分成彼此仿射同类的曲线所组成的仿射类. 原来把方程化成标准形状正好给出了这种分类，即二次曲线的仿射类是九个.（我们不预备详细说明为什么虚椭圆和一对虚的平行直线归于不同的仿射类. 严格地说，在平面上根本没有这样两种曲线. 谈到的已经是方程本身的代数性质了.）

同样地，二次曲面按其十七种标准方程的分类也是仿射分类.

让我们来给出一个应用二次曲面仿射分类的例子. 我们来证明，如果在空间中任意地选取这样三条直线 a, b, c：1)其中任何两条都不在一个平面上（即相错）和 2) 它们三条不同时平行于同一个平面，则空间中同时与所说三条直线 a, b 和 c 都相交的所有直线 d 的集合（图 72）形成整个一个单叶双曲面.

我们先指出这里谈到的直线 d 的集合是怎样的. 通过直线 a 的任意点 A 可以引包含直线 b 的平面 P 和包含直线 c 的平面 Q，这两个平面 P 和 Q 相交于唯一的直线 d，它通过直线 a 的点 A

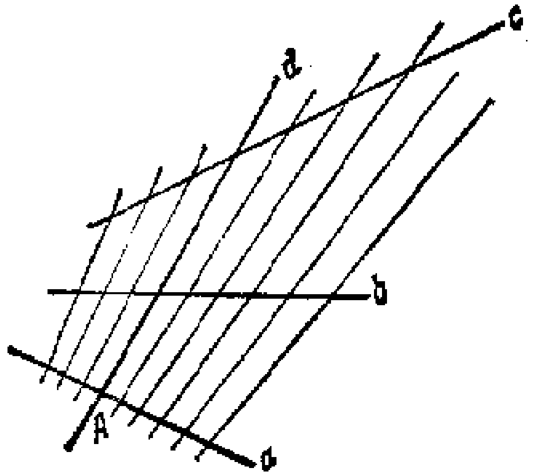


图 72

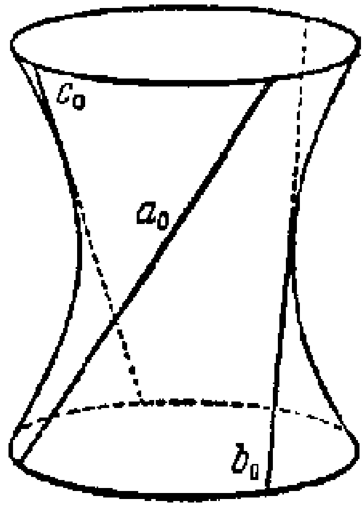


图 73

而且与直线 b 和 c 相交．通过直线 a 上的每个点都引这样的直线 d，我们就得到空间中与所有三条已知直线 a，b 和 c 都相交的直线 d 的集合．这个直线集合形成一个曲面．我们注意到，每一个给定的单叶双曲面都可以这样地得出，只要取同一族的三条不同的直母线 a_0，b_0，c_0 作为直线 a，b 和 c（图 73），而取另一族的所有母线作为直线 d．反过来，设给了空间中的任意三条两两相错的直线 a，b，c，它们不同时平行于同一个平面．那末可以证明，它们总是某个平行六面体的三条两两的没有公共点的棱所在的直线（图 74）．对于给定的直线 a，b，c 作出这样的平行六面体，对于某个单叶双曲面上同一族的三条母线 a_0，b_0，c_0 也这样办，然后我们来作把平行六面体 a_0，b_0，c_0 变成平行六面体 a，b，c 的空间仿射变换；它显然把这个双曲面变成我们所讨论的曲面．但是根据二次曲面的仿射分类，单叶双曲面的仿射像仍然是单叶双曲面．

3）应用于密集物质的连续变换的理论，例如弹性理论、流体理论、电磁场理论等等．所讨论的密集物质的很小的元素"几乎"是仿射地变换的．这就是所谓"局部的线性变换"（线性指的是一次式，而在下一段里我们将会看到，在解析几何里仿射变换的公式正是一次的）．在图 75 上就可以看出来，大的正方形网有了显著的变形，"扇子似地"散开了，等等．对不大的方块中的很小的正方形网来说，这一切就表现得很不显著了，而且这小正方形网"几乎"

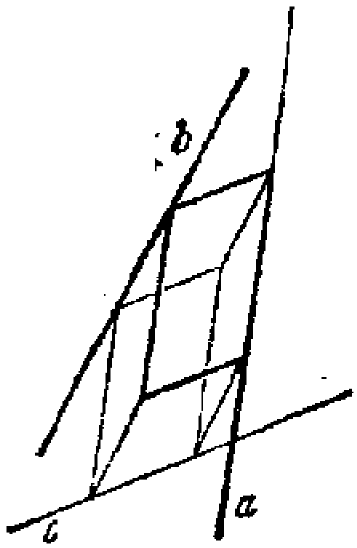


图 74

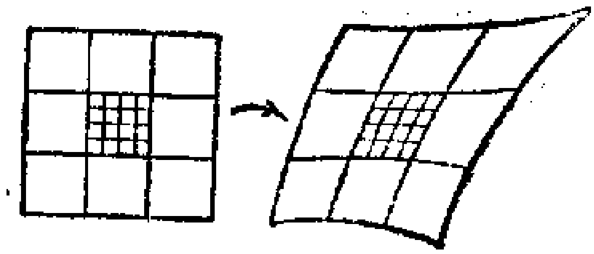


图 75

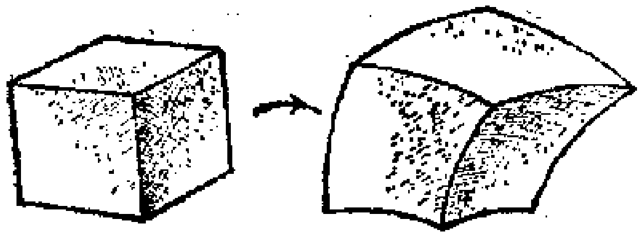


图 76

变成一个相等的平行四边形网了．在空间中也可以得到同样的图像(图 76)．由于空间的仿射变换可以化成一个移动和三个互相垂直的“压缩”，由此可知，物体的小块在弹性下所受的变化，首先是作为刚体的移动，此外还经受三个互相垂直的“压缩”．

仿射变换的公式　如果 Oe_1e_2 是经受仿射变换的标架，$O'e_1'e_2'$ 是它的像，而且新原点 O' 对于旧标架的坐标是 ξ，η，而向量 e_1' 和 e_2' 对于旧标架的坐标是 a_1，a_2 和 b_1，b_2，从图 77 中很容易看出，仿射变换的公式是

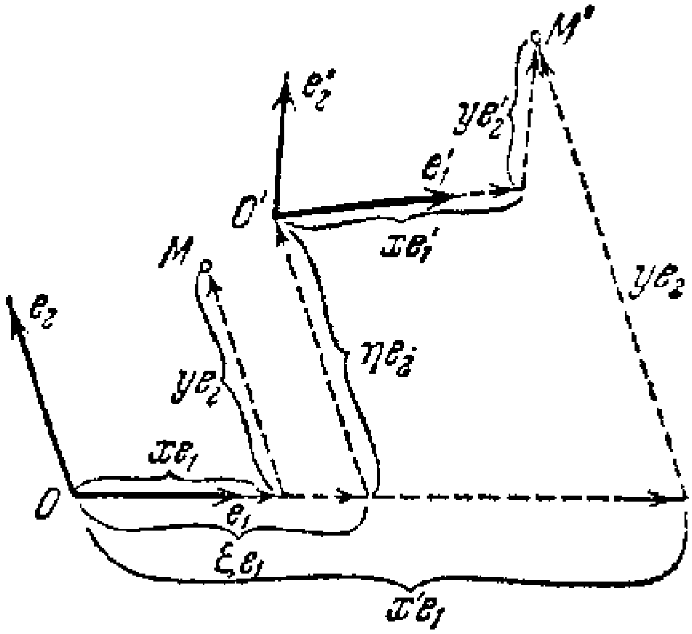


图 77

$$x' = a_1x + b_1y + \xi,$$
$$y' = a_2x + b_2y + \eta.$$

这公式的意思是这样：如果 x，y 是任意点 M 对于旧标架 Oe_1e_2 的

坐标，则由这两个公式表出的 x', y' 是这个点的像 M' 对于同一个旧标架的坐标。

实际上，设 $O\boldsymbol{e}_1\boldsymbol{e}_2$ 是被变换的标架，$O'\boldsymbol{e}_1'\boldsymbol{e}_2'$ 是它的像，M 是平面上任何一个被变换的点，M' 是它的像。那末根据仿射变换的定义，如果点 M 对于标架 $O\boldsymbol{e}_1\boldsymbol{e}_2$ 的坐标是 x, y, 则这点的像 M' 对于这标架的像 $O'\boldsymbol{e}_1'\boldsymbol{e}_2'$ 的坐标是同样的 x, y.

现在我们来考虑从被变换的标架的原点 O 引到点 M 的像 M' 的向量 $\boldsymbol{m}'$. 这时 $\boldsymbol{m}'=x'\boldsymbol{e}_1+y'\boldsymbol{e}_2$. 但是这个向量等于以下的向量和

$$\boldsymbol{m}'=\xi\boldsymbol{e}_1+\eta\boldsymbol{e}_2+x\boldsymbol{e}_1'+y\boldsymbol{e}_2',$$

而向量 $\boldsymbol{e}_1'$ 和 $\boldsymbol{e}_2'$ 又是

$$\boldsymbol{e}_1'=a_1\boldsymbol{e}_1+a_2\boldsymbol{e}_2,\quad \boldsymbol{e}_2'=b_1\boldsymbol{e}_1+b_2\boldsymbol{e}_2,$$

因此，
$$\boldsymbol{m}'=\xi\boldsymbol{e}_1+\eta\boldsymbol{e}_2+a_1x\boldsymbol{e}_1+a_2x\boldsymbol{e}_2+b_1y\boldsymbol{e}_1+b_2y\boldsymbol{e}_2,$$

或者
$$\boldsymbol{m}'=(a_1x+b_1y+\xi)\boldsymbol{e}_1+(a_2x+b_2y+\eta)\boldsymbol{e}_2.$$

比较 $\boldsymbol{m}'$ 的这个得到的表达式和上面的第一个表达式，我们得到

$$\left.\begin{aligned}x'&=a_1x+b_1y+\xi,\\ y'&=a_2x+b_2y+\eta.\end{aligned}\right\}\tag{21}$$

可以证明，行列式

$$\Delta=\begin{vmatrix}a_1 & b_1\\ a_2 & b_2\end{vmatrix}=a_1b_2-a_2b_1$$

不等于零而等于这样两个平行四边形的面积的比值，第一个平行四边形是作在新标架的向量上的，第二个平行四边形是作在旧标架的向量上的。

同样地可以得到空间的公式

$$\left.\begin{aligned}x'&=a_1x+b_1y+c_1z+\xi,\\ y'&=a_2x+b_2y+c_2z+\eta,\\ z'&=a_3x+b_3y+c_3z+\zeta,\end{aligned}\right\}\tag{22}$$

这里 (ξ, η, ζ) 是变换成的标架 $O'\boldsymbol{e}_1'\boldsymbol{e}_2'\boldsymbol{e}_3'$ 的原点 O' 对于被变换的标架 $O\boldsymbol{e}_1\boldsymbol{e}_2\boldsymbol{e}_3$ 的坐标，(a_1, a_2, a_3), (b_1, b_2, b_3), (c_1, c_2, c_3) 是向量 $\boldsymbol{e}_1'$, $\boldsymbol{e}_2'$, $\boldsymbol{e}_3'$ 的坐标。

行列式[1)]

$$\Delta=\begin{vmatrix}a_1 & b_1 & c_1\\ a_2 & b_2 & c_2\\ a_3 & b_3 & c_3\end{vmatrix}=a_1b_2c_3+a_2b_3c_1+a_3b_1c_2-a_1b_3c_2-a_2b_1c_3-a_3b_2c_1$$

不等于零而等于这样两个平行六面体的体积的比值，其中第一个是作在新标架的向量上的，第二个是作在旧标架的向量上的.

正交变换 平面作为刚体沿着自身的移动，以及这种移动加上对平面上某条直线的反射，都叫做平面的正交变换. 空间作为刚体的移动，以及这种移动加上空间对某个平面的反射，都叫做空间的正交变换. 明显地，正交变换是这样的仿射变换，它不改变标架的“度量”，即它只是经受移动，或者经受移动加上反射.

我们利用直角坐标来研究正交变换，即这时原来标架的向量互相垂直而且有长度等于绝对尺度单位，经过正交变换，标架的向量依然互相垂直(即它们的数量乘积等于零)而且它们的长度还等于1. 所以[参看第231页公式(14)]在平面的情形我们有

$$a_1b_1+a_2b_2=0,\quad a_1^2+a_2^2=1,\quad b_1^2+b_2^2=1, \tag{21'}$$

而在空间的情形

$$\begin{aligned}a_1b_1+a_2b_2+a_3b_3=0,\quad & a_1^2+a_2^2+a_3^2=1,\\ a_1c_1+a_2c_2+a_3c_3=0,\quad & b_1^2+b_2^2+b_3^2=1,\\ b_1c_1+b_2c_2+b_3c_3=0,\quad & c_1^2+c_2^2+c_3^2=1.\end{aligned} \tag{22'}$$

所以，如果原来的标架是直角标架，则公式(21)在而且只在正交条件(21')成立时才给出平面的正交变换，而公式(22)在而且只在正交条件(22')成立时才给出空间的正交变换，可以证明，如果 $\Delta>0$，则这是移动，而如果 $\Delta<0$，则这是移动加上反射.

§12. 不变量[2)]理论

不变量观念. 带两个变数的二次方程的不变量 在上一个卅

1) 关于行列式请参看第三卷第十六章.

2) 拉丁原文 invarians，意思是不变.

纪的后半还引入了一个重要的新的概念——不变量的概念.

例如我们考虑带两个变数的二次多项式

$$Ax^2+2Bxy+Cy^2+2Dx+2Ey+F. \tag{23}$$

如果把 x, y 看做直角坐标而且作从它们变到新直角坐标轴的变换，则在把(23)式中的 x, y 用它们以新坐标 x', y' 表达的式子代入以后，经过脱括弧和归并同类项，我们得到一个具有另外一些系数的变换成的新多项式

$$A'x'^2+2B'x'y'+C'y'^2+2D'x'+2E'y'+F'. \tag{24}$$

原来，存在着由系数组成的一些式子，它们的数值经过变换并不改变，虽然这些系数本身却改变了. 这种由 A', B', C', D', E', F' 组成的式子，它的数值就与它们是由 A, B, C, D, E, F 组成时相同.

这种式子叫做多项式(23)对于正交变换群的不变量(也就是对于从一组直角坐标 x, y 到任何另一组直角坐标 x', y' 的变换说的不变量).

下面就是这样的一些不变量:

$$I_1=A+C,$$

$$I_2=\begin{vmatrix} A & B \\ B & C \end{vmatrix}=AC-B^2,$$

$$I_3=\begin{vmatrix} A & B & D \\ B & C & E \\ D & E & F \end{vmatrix}=ACF+2BDE-AE^2-CD^2-FB^2,$$

即
$$A+C=A'+C',\quad AC-B^2=A'C'-B'^2,$$

$$ACF+2BDE-AE^2-CD^2-FB^2$$
$$=A'C'F'+2B'D'E'-A'E'^2-C'D'^2+F'B'^2.$$

可以证明这样一个重要的定理：多项式(23)的任何一个正交不变量都可以通过这三个基本不变量来表达.

如果让多项式(23)等于零，则我们就得到一条二次曲线的方程. 与这曲线本身有关而与这曲线在平面上的位置无关的每一个量，显然不依赖于它的方程是在那样的坐标里写出的，所以如果它

可以通过系数来表达，则这个表达式就是多项式(23)的正交不变量，因此，根据上面所说的定理，它就可以通过这三个基本的不变量来表达。再有，因为在用任意不是零的已知数 t 去乘所说方程的所有六个系数时，由这方程所表示的曲线依然是原来的，所以通过 I_1, I_2, I_3 来表达每一个只与曲线本身有关的量的式子，一定应该是这样的，当这式子中的 A, B, C, D, E, F 乘上 t 时，t 在这式子中正好能消去。因而可以说，所说的式子对于 A, B, C, D, E, F 应该是齐次的零次式。

我们用例子来验证这一点。譬如设方程

$$Ax^2+2Bxy+Cy^2+2Dx+2Ey+F=0$$

表示一个椭圆。因为这个方程完全决定了这个椭圆，所以利用这个方程，即利用它的系数，可以算出与这个椭圆有关的所有基本的量。例如可以算出它的半轴 a 和 b，即可以用系数来表示这些半轴。这些式子都是不变量，因此就是某些通过 I_1, I_2, I_3 来表达的式子。使用化方程为标准形状的方法和作一些进一步的计算，可以实地得出通过 I_1, I_2, I_3 来表达半轴的下列(十分复杂的)式子：

$$\sqrt{\frac{2|I_3|}{|I_2|\,|I_1\pm\sqrt{I_1^2-4I_2}|}}.$$

从以上所说的知道，由于不变量 I_1, I_2, I_3 虽然是系数的齐次式，但是并不是零次式，所以它们本身并无几何意义——它们只是代数的对象。

可以证明，表达式

$$K_1=\begin{vmatrix}A & D\\ D & F\end{vmatrix}+\begin{vmatrix}C & E\\ E & F\end{vmatrix}=AF-D^2+CF-E^2$$

虽然在平行移轴下会改变，但是在已知直角坐标轴的单纯的旋转下却不改变，这是所谓半不变量。

为了举例说明不变量和半不变量的应用，我们列出下面一个表，只要计算 I_1, I_2, I_3 和 K_1，这个表直接告诉我们如何从方程来决定它所表示的二次曲线的仿射类：

类的示志	名　称	归范方程	标准方程
$I_2>0,\ I_1I_3<0$	椭圆		$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$
$I_2>0,\ I_1I_3>0$	虚椭圆		$\frac{x^2}{a^2}+\frac{y^2}{b^2}=-1$
$I_2>0,\ I_3=0$	点	$\lambda_1x^2+\lambda_2y^2+\frac{I_3}{I_2}=0$	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=0$
$I_2<0,\ I_3\neq0$	双曲线		$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$
$I_2<0,\ I_3=0$	一对相交的直线		$\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$
$I_2=0,\ I_3\neq0$	抛物线	$I_1x^2+2\sqrt{-\frac{I_3}{I_1}}y=0$	$x^2=2py$
$I_2=0,\ I_3=0,\ K_1<0$	一对平行的直线		$x^2=a^2$
$I_2=0,\ I_3=0,\ K_1>0$	一对虚的平行直线	$I_1x^2+\frac{K_1}{I_1}=0$	$x^2=-a^2$
$I_2=0,\ I_3=0,\ K_1=0$	一对重合的直线		$x^2=0$

在这个表里列出了二次曲线的方程将会化成九种标准形状中那一种的必要充分条件(I_1I_3 表示 I_1 和 I_3 的乘积).

例如设给了方程 $x^2-6xy+5y^2-2x+4y+3=0$. 我们有 $A=1,\ B=-3,\ C=5,\ D=-1,\ E=2,\ F=3$, 于是 $I_1=6$, $I_2=-4,\ I_3=-9$. 表中第四行的条件成立, $I_2<0,\ I_3\neq0$, 即这是双曲线. 它的半轴等于

$$\sqrt{\frac{2\cdot9}{4\cdot|6\pm\sqrt{36+16}|}}\approx0.57 \text{ 和 } 1.93.$$

“归范”方程(I), (II)和(III)的系数用不变量和半不变量来表示, 即

$$\lambda_1x''^2+\lambda_2y''^2+\frac{I_3}{I_2}=0, \tag{I}$$

$$I_1x''^2+2\sqrt{-\frac{I_3}{I_1}}y''=0, \tag{II}$$

$$I_1x''^2+\frac{K_1}{I_1}=0, \tag{III}$$

这里 λ_1 和 λ_2 是下列所谓特征方程的根:

$$\lambda^2 - I_1\lambda + I_2 = 0.$$

用公式(I)—(III)可以直接算出椭圆和双曲线的半轴 a 和 b，抛物线的参数 p 和平行直线之间的距离 $2a$. 关于半轴的公式前面已经写出过了. 参数 p 可以得出是等于

$$p = \sqrt{-\frac{I_3}{I_1^3}},\ \text{距离}\ 2a = 2\sqrt{-\frac{K_1}{I_1^2}}.$$

关于三维空间中的二次曲面，也可以引出完全相类似的不变量和半不变量理论，以及对应的用于决定仿射类和归范方程系数的公式的表.

必须指出，以上的全部叙述只是说明了在解析几何学里所讨论的那些不变量的意义和价值. 然而不变量概念本身却还有非常重大的价值.

一种被研究的对象对于它的一种被考虑的变换的所谓不变量，是指与这个对象有关而在这种变换下不变的每一个量(数量，向量或其他量). 在已经讨论过的问题里，对象是带两个变数的二次多项式(其实是它的系数)，变换是在从一组直角坐标过渡到另一组直角坐标而得到的多项式的变换.

另一个例子：对象是在已知温度下的已知气体的已知质量. 变换是这个质量气体的体积或压力的改变. 按照玻义耳-马利奥特定律，体积和压力的乘积就是一个不变量. 可以说空间中的线段的长度或者角的角度是空间的移动群的不变量，点分线段的比值或体积的比值是空间的仿射变换群的不变量，等等.

各种不变量在物理学中有特别重要的价值.

§13. 射影几何

透视投射 古代的画家就已开始研究透视的规律了. 这之所以必须，是因为物体是透视投射到眼睛的网膜上而被人们看到的，这时物体的形状和相互位置要遭受一些异样的歪曲. 举例说，远看过去的电线杆子好像愈来愈小也愈来愈密，平行的铁轨看来像

要逐渐并在一起似的，等等，我们现在不预备讨论空间的透视，即空间对象到平面上的透视投射的性质，而只讨论平面到平面的透视投射的性质.

设有画片(例如电影胶片)P，幕P'和在它们之间的透镜S(图 78)．于是如果画片是透明的而且从后面照亮了它(如果它不透明，则从前面，即从透镜所在的一面照亮它)，则画片上被照亮的各点发出的光线束，经过透镜聚集，它们还以点的形状被投射到幕P'上．我们认为所发生的情况是这样：就好像画片P上的点被投射到幕P'上是从通过透镜的光学中心的直线而进行的．

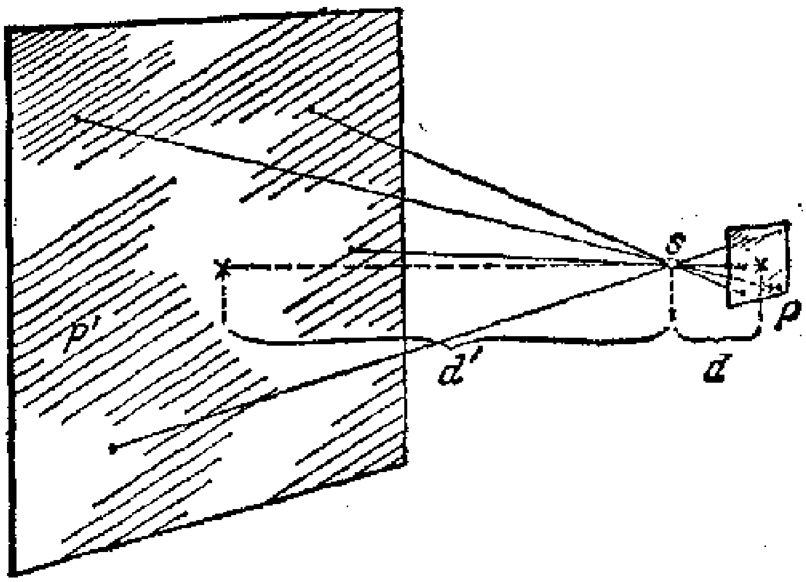


图 78

假如平面P和P'是平行的，事情是非常简单的．在这种情形里，在平面P'上得到的显然是与平面P上所有图形相似的绘像．这绘像比原物大了些还是小了些，就看比值$d':d$(这里d和d'分别是从透镜中心到平面P和P'的距离)小于1还是大于1．

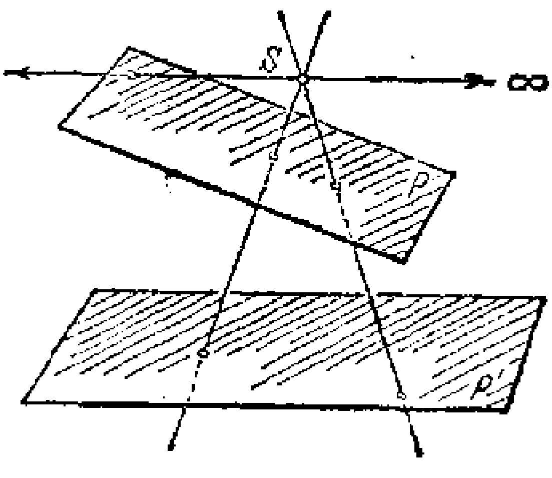


图 79

特别复杂的是平面P和P'不平行的情形(图 79)．在

这种情形里，在通过 S 的投射下，不但图形的大小改变了，而且它们的形状也歪曲了，这时平行直线可以被投射成不平行的；点分线段的比值可以改变，等等．一般地说，甚至在任意仿射变换下都不变的一些关系这时也会改变．

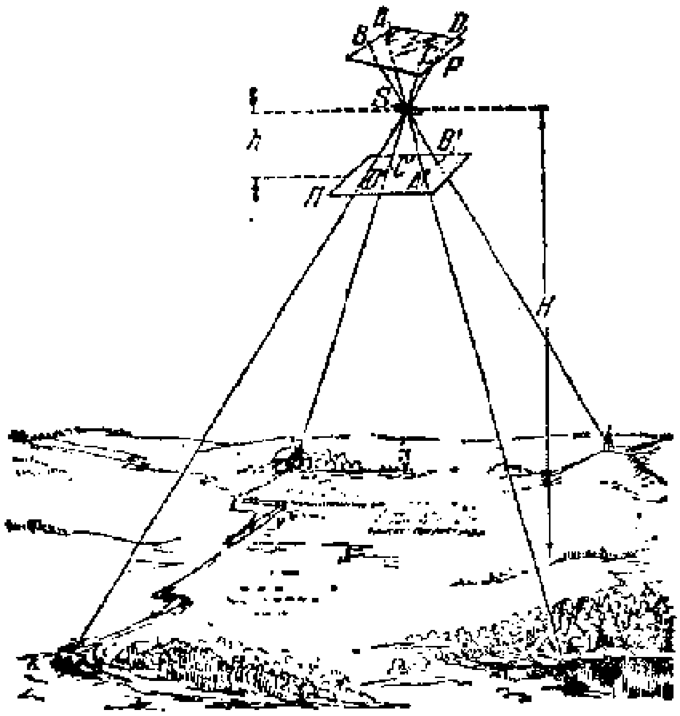


图 80 a

例如在航空摄影时就有这种投射．飞机在飞行时要动荡，因而牢固地装在飞机上的摄影机（图 80 a）一般地说并非正好铅垂地从上面对着物体，而是在拍摄时位置时常有些倾斜，因而得到的只是地面的变了形的绘像（我们假定地面是平的）．

如何矫正这个绘像呢？为此必须研究一个平面 P 到另一个平面 Π 的投射的性质．一般地说这两个平面不平行，而且投射是由通过既不在平面 P 上又不在平面 Π 上的点 S 的直线来实现的．这种投射叫做**透视投射**．

我们以后要证明下面的重要定理．

定理　如果有从平面 P 到平面 Π 的这样两个透视投射，在这两个投射下，平面 P 上"一般位置"的四个点 A, B, C, D（即其中任何三个点都不在一条直线上）都被投射成平面 Π 上同样的四个点 A', B', C', D'，则平面 P 的任何点在这两个投射下也被投射成平面 Π 的同一个点．

换句话说，如果知道了被投射的图形中任何四个一般位置的点在一个透视投射下变成什么样的点，这个投射的结果就完全决定了．

这就是所谓射影映射理论中的唯一性定理或者说是平面投射的基本定理.

平面投射的基本定理在航空摄影中的应用 我们要来说明这个定理如何给出矫正摄影绘像的适当方法.

如果设想在作航空摄影时有一个水平幕 Π 放在这样的位置：透镜的中心 S 在它之上高度为 h 处（图 80 a），则通过 S 投射到这幕上以便由照相底片 D 摄下的绘像显然未经歪曲，而是以比例尺 $h:H$ 与水平地区相似的，这里 H 是在摄影时飞机离地面的高度. 为了矫正在底片 P 上得到的绘像而且使它转为正确的，人们这样做：把底片插入装在特殊架子上的投射器里，投射器利用调节螺丝在架子上可以调整离幕布 Π 的距离，而且可以随意改变它的方位.

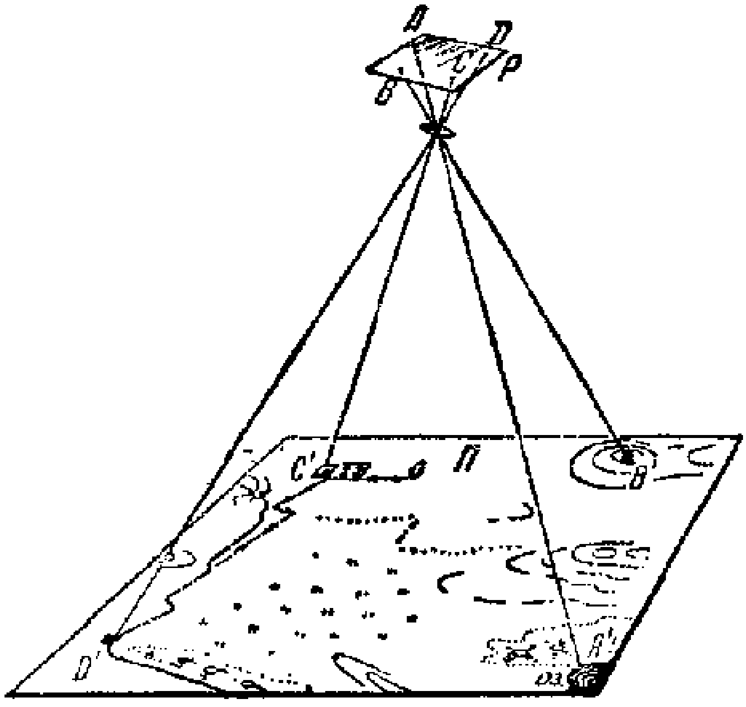


图 80 b

把测量地面而得到的一张地形图（不必是详细的，虽然在照相底片上有着我们所感兴趣的全部细节，但是这些在地图上却并非必需的）装在幕 Π 上（图 80b）. 在这装在幕 Π 上的地图上选取四个点 A', B', C', D'，这些点易于在底片上找到（例如交叉道路的一角，房屋的一角，等等），而且在底片 P 上受到感光的对应的点 A, B, C, D 处用针刺成小孔. 在底片 P 后面装一个投射用的灯，因此底片就通过它所插入的投射器的透镜 S 而被投射到幕 Π 上. 运用调节螺丝，使得刺穿的小孔正好对准装在幕 Π 上的地图上对应的点 A', B', C', D'. 做到了这一步以后，把地图换成带照相正

片的夹子，不要更动装置，在正片上就能得到从飞机所照的底片 P 到幕 Π 上的投影.

根据上面所说的定理，在正片上得到所摄地区的正确的（即与地面相似的）而不是歪曲的地图.

现在让我们来叙述为证明所说基本定理所必需的理论.

射影平面 空间中通过已知点 S 的全体直线和平面的集合，叫做具有中心 S 的投射直线和平面丛. 假如让不过丛中心的平面 P 与丛相交，则对应于平面 P 的每个点，有丛中与平面 P 相交于这个点的直线，对应于平面 P 的每条直线，有丛中与平面 P 相交于这条直线的平面. 然而这样还不能建立从丛中直线和平面的集合到平面 P 上点和直线的集合的一一映射. 问题在于：丛中平行于平面 P 的直线和平面，因为不与平面 P 相交，所以在上述意义下不对应于平面 P 上的任何点和直线. 然而大家约定说：丛中的这些直线还与平面 P 相交，只是相交于平面 P 上处在某些方向的非本义的点（或者无穷远的点）；丛中的这个平面也还与平面 P 相交，只是相交于平面 P 的非本义的直线（或者无穷远的直线）. 增添了这些非本义点和非本义直线以后的平面 P，叫做增广平面或者射影平面. 我们把它记做 P^*. 从 S 中的直线和平面的集合现在可以一一地映射到这个射影平面 P^* 的点（本义的和非本义的）和直线（本义的和非本义的）的集合了.

此外，大家还约定，假如丛中对应的直线处在丛中对应的平面上，就说射影平面 P^* 的点（本义的或非本义的）处在直线（本义的或非本义的）上. 从这个观点看来，射影平面的任何两条直线都相交（于本义的点或非本义的点），因为丛中任何两个平面总相交于一条直线. 由此顺便推出，非本义的直线正是全部非本义的点的集合.

事情的实质是这样：平面在增添了它的非本义的元素以后，我们就可以利用这个平面作为截平面，来研究通过一个点的全体直线和平面组成的丛.

射影映射；基本定理 射影映射是指射影平面 P^* 到另一个射

影平面 $P^{*\prime}$ 的这种映射(平面 $P^{*\prime}$ 可以与平面 P^* 重合，那时就说是平面 P^* 的射影变换)，它首先是点的一个一一映射，其次它把平面 P^* 的共直线的点集合变成平面 $P^{*\prime}$ 的也共直线的点集合，反过来也这样.(这时总认为点和直线既指本义的点和直线，也指非本义的点和直线.)

明显地，从同一个平面 P^* 到某个平面 Π^* 的任何两个透视投影可以彼此经过射影变换而得到.

事实上：1° 它们的点(本义的和非本义的)都与射影平面的点(本义的和非本义的)一一对应，因此它们也彼此一一对应，2° 第一个投影的共直线的点对应于平面 P^* 的共直线的点，因而就对应于第二个投影的共直线的点，反过来也这样. 所以前面所说的透视理论中的定理是关于射影变换的下列定理的直接推论：如果在射影平面 Π^* 的某个射影变换下，它的四个一般位置的点 A, B, C, D 留在原地，则它的所有的点都留在原地.

我们提出利用所谓麦比乌斯网来证明这个定理的想法.

我们注意到：1) 如果在一个射影变换下有两个点留在原地，则通过它们的直线变成它自己；2) 如果有两条直线变成它们自己，则它们的交点留在原地，所以由于平面 Π^* 的点 A, B, C, D 留在原地，顺次地点 E, F, G, H, K, L 等等也留在原地(图81). 连接已经得到的点，还可以继续作出这样的点. 这就是所谓麦比乌斯网. 继续这样的作图，可以使它细密直到无穷. 可以证

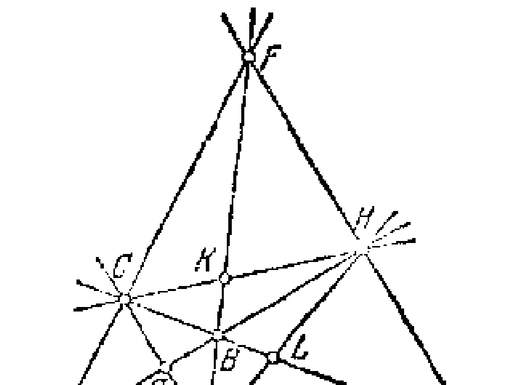


图 81

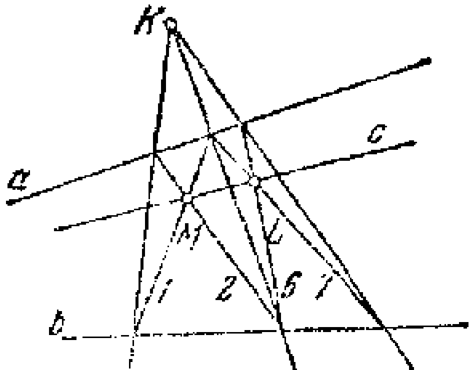


图 82

明，这个网的全部交叉点到处稠密地盖在整个平面上．所以，如果再假定射影变换的连续性（这虽然可以从其定义推出，但是不很容易），那末可以证明，假如在平面 Π^* 的射影变换下点 A, B, C, D 留在原地，则平面 Π^* 的所有的点都留在原地．

射影几何　所谓二维的射影几何，是指关于射影平面上图形的射影性质的定理全体．射影平面就是增添了非本义元素的普通平面．射影性质是指在任何射影变换下不变的那些性质．

这里是射影几何中问题的例子．给了两条直线 a 和 b 和一个点 M（图 82）．不利用直线 a 和 b 的交点本身（例如当这个交点非常远时，可以是必需这么办的），作出通过点 M 和这个交点的直线 c．通过点 M 引两条割线 1 和 2，然后通过它们与直线 a 和 b 的交点作直线 3 和 4，我们得到了点 K．通过点 K 再引割线 5 并且作出直线 6 和 7，于是可以证明，通过直线 6 和 7 的交点 L 和点 M 的直线 c 就是所求的直线．

从圆锥截线的理论得知（图 83），椭圆、双曲线和抛物线互相成为透视投影，并且它们都是圆的透视投影．

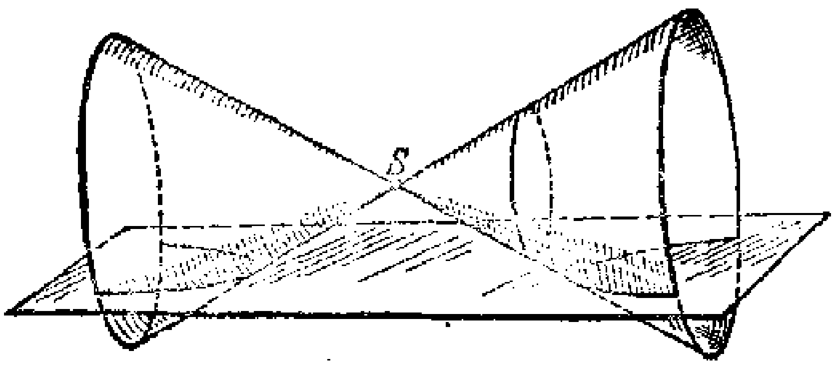


图　83

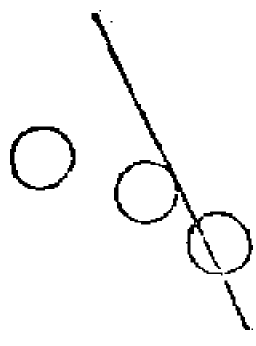


图　84

如果把透视投射看做射影平面 P^* 和 $P^{*\prime}$ 彼此间的射影映射，则在把这两个平面合并成一个以后，我们得知所有椭圆、双曲线和抛物线都是圆周经过射影变换的结果．不同的是，对于圆周所经受的射影变换说，假如它把不与圆周相交的直线变成无穷远直线，则得到椭圆；假如它把与圆周相切的直线变成无穷远直线，则得到抛物线；假如它把与圆周相交的直线变成无穷远直线，则得到双曲线（图 84）．

射影变换的公式 如果在平面上取普通的笛卡儿坐标，则可以证明，平面的射影变换的公式有形状

$$x'=\frac{a_1x+b_1y+c_1}{a_3x+b_3y+c_3},\quad y'=\frac{a_2x+b_2y+c_2}{a_3x+b_3y+c_3},$$

这里行列式

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}\neq 0,$$

反过来也对.

如果对于某个点(x, y)说，分母成为零，则就认为它的像$(x'y')$是非本义点（无穷远点）. 方程

$$a_3x+b_3y+c_3=0$$

表示一条直线，它在所说的射影变换下变成非本义直线（无穷远直线）.

§14. 罗仑兹变换

从关于光的传播速度的恒定性导出沿直线运动的罗仑兹变换的公式和平面上运动的罗仑兹变换的公式 在十九世纪末，在物理学中发现了一个基本的矛盾. 迈开尔逊作了他的著名实验，在地球沿着它的轨道绕日运动（地球的速度大约是每秒30公里）的方向和垂直于这个方向测量光速（大约是每秒300,000公里），无可争辩地证明了：自然界的所有物体在运动时都要在运动的方向收缩，即使在真空中也是如此. 荷兰物理学家罗仑兹详细地研究了这个收缩的理论. 原来物体的运动速度愈接近光在真空中的速度，这种收缩就愈厉害，并且当物体的运动速度等于光速时，它就无限制地收缩. 罗仑兹写出了这个收缩的公式. 然而不久物理学家爱因斯坦以完全另外一种观点提出了这同一个问题，这个观点是庞加来所曾经接触到的. 爱因斯坦这样地进行论证：如果认为对于光的传播，也像对于普通的物体运动一样，伽利略的速度相加定律成立，则光的速度$c'=c+v$，这里v是迎着光的传播方向运动

的观察者的速度，而 c 则是对于不动的观察者说的光速. 然而从迈开尔逊的实验得知 $c'=c$. 等式 $c'=c+v$ 以下列变换为根据：

$$\begin{aligned} x' &= x+v_x t, \\ t' &= t, \end{aligned} \tag{25}$$

这式子联系了一个点对于某个坐标系 I 的坐标 x 与它对于坐标系 II 的坐标 x'，坐标系 II 的轴平行于坐标系 I 的轴，而且坐标系 II 向着 Ox 轴的方向以速度 v_x 相对于坐标系 I 而运动. 很明显，正像爱因斯坦所说的，这些公式应该是不变的.

可以证明(例如像亚历山大洛夫最近所做的)，只要从两个坐标系 x, y, z, t 和 x', y', z', t' 中的一个光速的等式，就可以知道从坐标 x, y, z, t 到坐标 x', y', z', t' 的变换公式是线性的和齐次的，即有形状

$$\begin{aligned} x' &= a_1x+b_1y+c_1z+d_1t, \\ y' &= a_2x+b_2y+c_2z+d_2t, \\ z' &= a_3x+b_3y+c_3z+d_3t, \\ t' &= a_4x+b_4y+c_4z+d_4t. \end{aligned} \tag{26}$$

从别的想法可以证明，这个变换公式的行列式[1)]等于 1.

如果有一个点在坐标系 I 里以光速 c 向着任意已知方向作等速直线运动，则 $x=v_x t$, $y=v_y t$, $z=v_z t$, 和 $v_x^2+v_y^2+v_z^2=c^2$, 于是

$$x^2+y^2+z^2-c^2t^2=0. \tag{27}$$

但是按照迈开尔逊的实验，这个点在坐标系 II 里应该以同样的光速 c 而运动，因此还应该有

$$x'^2+y'^2+z'^2-c^2t'^2=0.$$

因此，公式(26)并不是任意的具有等于 1 的行列式的齐次线性变换，而且还有这样的条件，如果 x, y, z, t 满足方程

$$x^2+y^2+z^2-c^2t^2=0,$$

则它们经过变换而得到的 x', y', z', t' 也满足这个方程. 这种变换(26)叫做**罗仑兹变换**.

我们先来讨论当点沿着 Ox 轴运动的最简单的情形. 在这情

1) 参看第三卷第十六章.

形里公式(26)有形状

$$x'=a_1x+d_1t,\qquad t'=a_2x+d_2t, \tag{26'}$$

而方程(27)变成

$$x^2-c^2t^2=0. \tag{27'}$$

引入记号 $ct=u$，于是公式(26')和方程(27')成为

$$x'=a_1x+\frac{d_1}{c}u,\qquad u'=a_2cx+\frac{d_2c}{c}u \tag{26_1}$$

和

$$x^2-u^2=0. \tag{27_1}$$

我们来求另一种形状的公式(26_1). 我们把 x 和 u 看做平面上的直角坐标，即我们从几何上来考虑问题，我们并且认为公式(26_1)是平面 Oxu 的仿射变换的公式（上面说过，它的行列式等于1），我们把这个变换记做 L. 如果按照我们的假定，从 $x^2-u^2=0$ 推出 $x'^2-u'^2=0$，则这个变换把一对相交的直线

$$x^2-u^2=0$$

变成自己. 因此，变换 L 是沿着这两条直线作具有相同系数 τ 的压缩和伸展的联合.

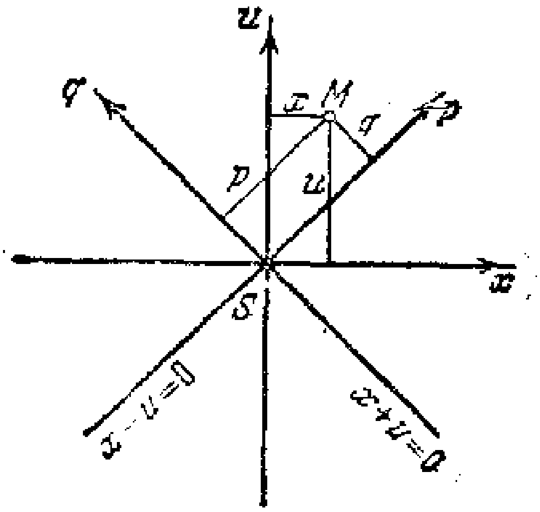


图 85

从图 85 我们得到

$$x=\frac{p}{\sqrt{2}}-\frac{q}{\sqrt{2}},$$

$$u=\frac{p}{\sqrt{2}}+\frac{q}{\sqrt{2}}.$$

而因为经过变换 L 以后，p 和 q 变成 $p'=\frac{p}{\tau}$ 和 $q'=q\tau$，所以

$$x'\sqrt{2}=\frac{p}{\tau}-q\tau,$$

$$u'\sqrt{2}=\frac{p}{\tau}+q\tau.$$

从第一对方程求出由 x 和 u 表达的 p 和 q，代入第二对方程，我们得到

$$x'=\frac{x-\frac{\tau^2-1}{\tau^2+1}ct}{\frac{2\tau}{\tau^2+1}},\quad t'=\frac{t-\frac{1}{c}\frac{\tau^2-1}{\tau^2+1}x}{\frac{2\tau}{\tau^2+1}},$$

或者在这时令 $\frac{\tau^2-1}{\tau^2+1}c=v$，我们得出

$$x'=\frac{x-vt}{\sqrt{1-\left(\frac{v}{c}\right)^2}},\quad t'=\frac{t-\frac{vx}{c^2}}{\sqrt{1-\left(\frac{v}{c}\right)^2}},$$

这就是著名的罗仑兹公式.

如果现在特别取 $x=0$，即考虑坐标系 I 的原点的运动，则我们得到

$$x'=\frac{-vt}{\sqrt{1-\left(\frac{v}{c}\right)^2}},\quad t'=\frac{t}{\sqrt{1-\left(\frac{v}{c}\right)^2}},$$

或者 $x'=-vt'$. 由此可见，v 是坐标系 II 相对于坐标系 I 的运动速度.

例如设给了 Ox 轴上的在坐标系 I 里具有坐标 x_1 和 x_2 的两个点，在坐标系 I 里它们之间的距离是 $r=|x_1-x_2|$. 我们来看一下，对于处在第二个坐标系里的观察者来说，它们之间的距离是怎样的. 我们有

$$x_1'=\frac{x_1-vt}{\sqrt{1-\left(\frac{v}{c}\right)^2}},\quad x_2'=\frac{x_2-vt}{\sqrt{1-\left(\frac{v}{c}\right)^2}},$$

于是

$$r'=|x_1'-x_2'|=\frac{|x_1-x_2|}{\sqrt{1-\left(\frac{v}{c}\right)^2}}.$$

因子 $\sqrt{1-\left(\frac{v}{c}\right)^2}$ 显然正是罗仑兹收缩的系数. 因为 c 非常大，所以对于不很大的 v，这个系数非常接近于1，因而收缩也就不显著. 然而一些基本的微粒，例如电子和正电子，常常以类似于光速的速度而运动，因而在研究它们的运动时就不能不考虑到这

个事实，也就是不能不考虑到所谓相对论的效果.

现在让我们来研究下列比较复杂的情形：当点在平面 Oxy 上运动的情形. 对于这种情形，变换(26)有形状

$$\begin{aligned} x' &= a_1x + b_1y + d_1t, \\ y' &= a_2x + b_2y + d_2t, \\ t' &= a_3x + b_3y + d_3t, \end{aligned} \qquad (26'')$$

这里

$$\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix} = 1,$$

而方程(27)则是

$$x^2 + y^2 - c^2t^2 = 0. \qquad (27'')$$

这是关于在平面 Oxy 上的运动的罗仑兹公式.

再令 $ct = v$. 于是变换(26″)就可以写成：

$$\begin{aligned} x' &= a_1x + b_1y + \frac{d_1}{c}u, \\ y' &= a_2x + b_2y + \frac{d_2}{c}u, \\ u' &= a_3cx + b_3cy + \frac{d_3c}{c}u, \end{aligned} \qquad (26_2)$$

并且它的行列式还等于 1，而方程(27″)则有更简单的形状

$$x^2 + y^2 - u^2 = 0. \qquad (27_2)$$

我们把 x, y, u 看做普通三维空间中点的笛卡儿直角坐标，而且把公式(26_2)看做这个空间的仿射变换的公式. 方程(27_2)表示一个顶角是 90° 的正圆锥 K (图 86).

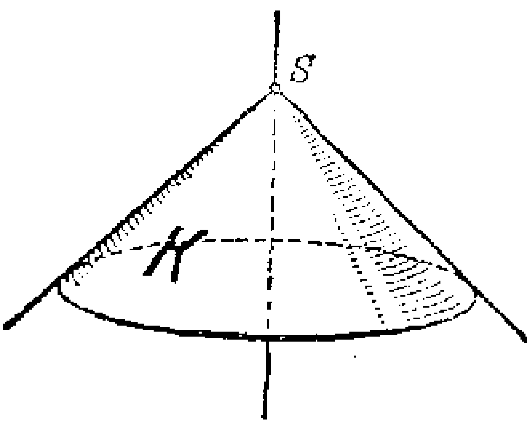


图 86

从罗仑兹变换的这个几何解释(因为我们在这里根本就把 $u=ct$ 看做空间的坐标)看来，平面上的运动不是别的，正是这个空间的把锥面 K 变成自己的所

有等积仿射变换(即不改变体积的仿射变换).

让我们来讨论某些这种特殊的罗仑兹变换.

1. 明显地，空间作为刚体绕锥面 K 的轴转一个角 ω 的任何简单的旋转，是空间的把锥面 K 变成自己的等积仿射变换，即是一些特殊的罗仑兹变换．我们把它记做 ω.

2. 空间对通过锥面 K 的轴的任何平面的反射，显然也是这种罗仑兹变换．我们把它记做 π.

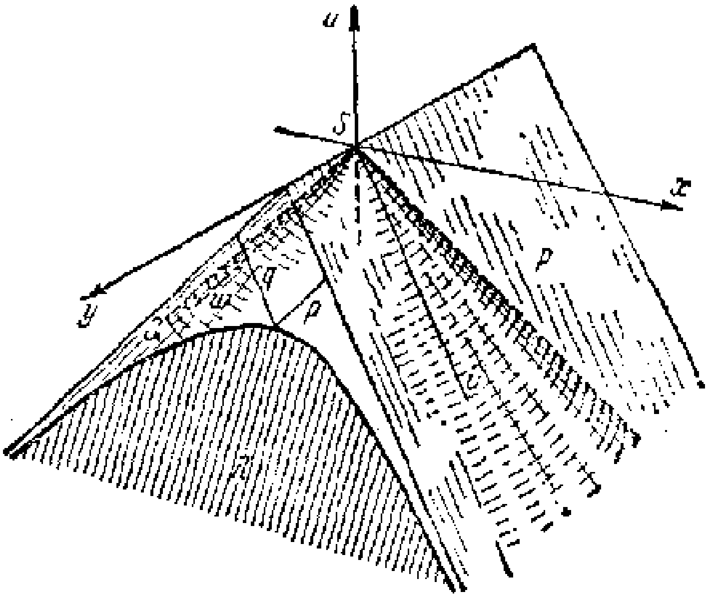


图 87

3. 最后，让我们来看一下以下的变换(图 87)．设 v 和 w 是锥面 K 的某一对对面的母线，P 和 Q 是沿这两条母线与锥面相切的平面．这两个平面互相垂直．我们作空间向着平面 P 的压缩和它背着平面 Q 而具有同一个系数的伸展，或者反之．例如我们把空间向着平面 P 压缩到三分之一，然后背着平面 Q 把它伸展到三倍．这种空间变换显然是仿射变换而且是保留所有体积的大小的．我们把它记做 L．让我们来证明，这个变换把锥面 K 变成自己．由于锥面 K 以 u 轴作为旋转轴，总可以把整个图形旋转到使母线 v 和 w 处在例如平面 Sxu 上．所以只要对于这种特别情形引出证明就成了．

为了证明，我们使平行于平面 Sxu 的任意平面 R 与锥面 K 相交．这个平面有方程 $y=b$，这里 b 是常数．把这个值代入锥面 K 的方程，我们得到

$$x^2-u^2=-b^2,$$

这是双曲线的方程，对于它说，平面 R 与平面 P 和 Q 的交线正是

渐近线．但是因为刻画这个双曲线的点的是：它到渐近线(即到平面 P 和 Q)的距离 p 和 q 的乘积是常数，在变换 L 下，这种双曲线上的每一个点都留在同一个双曲线上，因而双曲线就变成了自己，但是整个锥面 K 就由这种双曲线组成，所以在空间变换 L 下锥面 K 变成了自己．因此这个变换 L 也是罗仑兹变换．

因为在仿射变换下直线变成直线，而且相交直线变成相交直线，所以直线丛 S 在任何罗仑兹变换下一对一地变成自己．此外，因为在空间的仿射变换下每一个平面都变成一个平面，所以在丛 S 到自己的这种变换下得到的是这个丛的射影变换，如果用一个垂直于锥面 K 的轴的平面 Π 与这个丛相交，而且认为这个平面完全不参与空间的所说罗仑兹变换，再把这个平面增补成为射影平面 Π^*，来考虑丛 S 的直线与平面 Π^* 的交点，则变换丛 S 的罗仑兹变换，顺带给出平面 Π^* 的一个射影变换 Λ，它把由平面 Π^* 与锥面 K 的内部相交而成的圆形 α 变成自己．因此，为了研究罗仑兹变换的性质，最简单的是考虑它的所谓把圆 α 变成自己的射影变换 Λ．

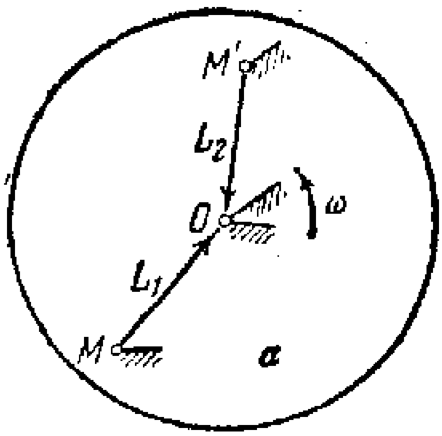


图 88

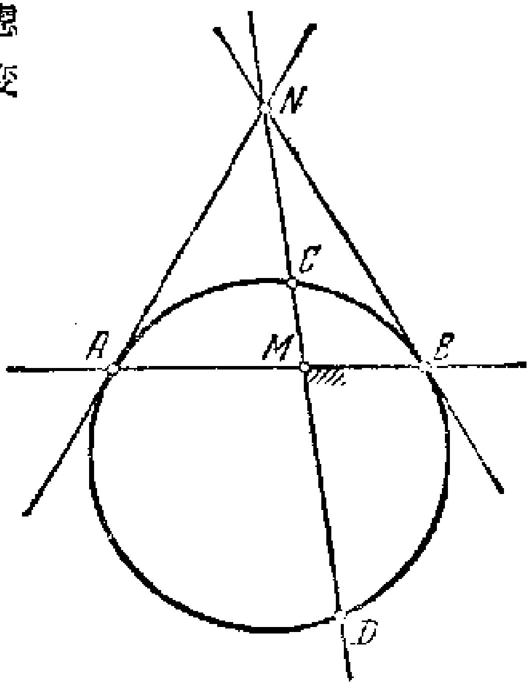


图 89

把圆变成自己的射影变换 我们把平面 Π^* 上的一个点，由这点引出的一条半直线和以这半直线所在的直线为界的一个半平面，合称为平面 Π^* 的"标架"(不要与 § 11 中的坐标架相混)．我

们来证明，如果取任意的两个标架 M 和 M'．它们的点都是圆 α 的内点，则可以用变换 L, ω 和 π 把其中第一个标架变成第二个标架．为此只要作变换 $\Lambda=L_1\cdot\omega\cdot L_2^{-1}$(或者变换 $\Lambda=L_1\cdot\omega\cdot\pi\cdot L_2^{-1}$)．变换 L_1 把第一个标架 M 变到圆 α 的中心 O，变换 ω 把它旋转到必要的位置，最后变换 L_2^{-1} 把它变到与第二个标架 M' 重合．

此外，让我们来证明，这种把已知标架 M 变成已知标架 M' 的变换 Λ 只有一个．为此我们先注意到，如果有两个不同的变换 Λ_1 和 Λ_2 都把标架 M 变成标架 M'，则变换 $\Lambda=\Lambda_1\Lambda_2^{-1}$ 就不是恒同变换而把标架 M 变成自己．因此只要能证明：假如有某个变换 Λ 把标架 M 变成自己，则它一定是恒同变换，即它保留圆 α 所在平面的全部的点在原地．

我们来证明这一点．设有一个变换 Λ 把标架 M 变成自己．那末它就把这个标架中的直线 AB 变成自己，但是因为它把圆 α 的圆周变成自己，所以它保留点 A 和 B 在原地或者交换它们的位置．然而因为它把标架中的半直线变成自己，后一情况不可能发生．在点 A 和 B 处引圆 α 的切线．它们都变成自己，因为假如这种切线变成割线 $A\overline{A}$，则反逆变换就要把圆周 α 的不同的点 A 和 $\overline{A}$ 变成同一个点 A．但是变换 Λ 是射影变换，因而是一对一的．一当变换 Λ 把这些切线变成自己，那末它们的交点 N 就留在原地，因此直线 MN 变成自己．再由于标架 M 中的半平面变成自己，我们像前面一样得知：点 C 和 D 不改变位置而留在原地．总之，在射影平面 Π^* 的所说射影变换 Λ 下，它的四个点 A, B, C, D 留在原地，而这四个点中任何三个点都不在一条直线上，因此，根据射影变换的基本定理，这个变换是射影变换．

以后在第十七章§5(第三卷)中将会证明，利用刚才引出的罗仑兹群的性质，很容易实现罗巴切夫斯基的平面几何，而如果对于点在空间中运动的一般情形讨论罗仑兹变换，则还可以实现罗巴切夫斯基的立体几何，因而也就证明了罗巴切夫斯基几何的无矛盾性．

我们看到，罗仑兹变换的理论，射影几何和透视理论，以及欧

几里得几何，都彼此有着紧密的联系。原来，与它们在很多方面有联系的还有复变数函数论里的所谓保角变换的理论，还解决了数学物理的很多重要问题，例如在受热的棒上温度分布的问题，关于空气对机翼的绕流问题，平面静电场理论的问题，弹性理论的平面问题以及很多别的问题。

结 束 语

解析几何是研究数学其他分支、力学、物理学和其他自然科学所十分必要的数学方法。所以不仅在大学里要学习它，而且在所有高等技术学校甚至某些中等技术学校里也要学习它。现在已经提出了把十分详细的解析几何原理放在中学课程里的问题。

各种坐标 我们看到，解析几何观念的重要部分是坐标方法和把方程看作联系这些坐标的关系的看法。除去笛卡儿坐标，还可以讨论各种别的坐标。譬如说，在平面上取定一个点 P（所谓极点）和从它引出的一条半直线（所谓极轴）以后，可以用这样两个数来决定一个点 M 的位置，从极点引到这个点的极半径的长度 ρ 和这个半径与极轴组成的角 ω（图 90）。

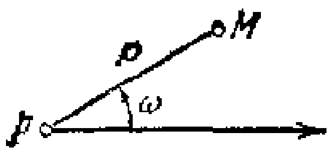


图 90

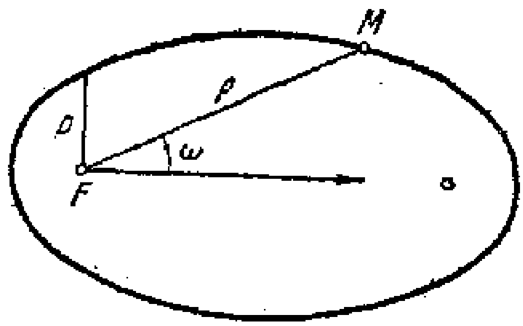


图 91

特别地，对于椭圆、双曲线或者抛物线，假如取焦点作为极点，而取从焦点沿着对称轴引向与最近的顶点相反一侧的半直线作为极轴（图 91），则这些曲线就有同一个方程

$$\rho=\frac{p}{1-\varepsilon\cos\omega},$$

这里 ε 是曲线的离心率，而 p 则是它们的所谓参数。这个方程在天

文学中有很大的价值.那是说,利用这个方程可从研讨如何从惯性定律和万有引力定律推导出行星是以椭圆轨道绕着太阳运行的.

大家都知道,地理坐标(经纬度)给定了点在球面上的位置.

同样地,就像在微分几何中所做的(参看第二卷第七章),可以在任何曲面上引入坐标网以及其他.

多维解析几何和无穷维解析几何.代数几何 解析几何到十九世纪时已经经历了这样长的发展道路,我们上面只是写出了大概的轮廓,而且它居然给出了那么多观念,看来它应已发展到了尽头吧,然而事实上却并非如此.就在现代,两个新的宽广的数学分支——所谓泛函分析和一般的代数几何还在迅速地发展着.固然,它们都只是一半地代表了古典解析几何的直接的延续:在泛函分析里有许多是分析,在代数几何里有不少函数论和拓扑学.

让我们来说明上一段话的意思.还在上一世纪中,人们已经开始讨论四维的和一般 n 维的解析几何,这就是说,开始研究这样的代数问题,它们是在二维和三维解析几何里所研究的代数问题的直接推广,即考虑有 4 个或 n 个未知数的情形.在十九世纪末,许多卓越的分析学家有了这样一种想法:为了分析和数学物理的目的,必须研讨无穷维的解析几何.

初看起来可以认为,假如 n 维空间甚至四维空间都像是不自然的数学上的虚构,那末无穷维空间就更是如此了.然而事实并不是这样.关于无穷维空间的论证,完全不是这样的困难.它们目前成为数学的一大分支——泛函分析(参看第三卷第十九章).在数学的这个部门里,苏联学者在最近做出了一些最重要的结果.

有趣的是,无穷维的解析几何有重要的实际应用而且在现代物理学里占有基本的地位.

至于说到代数几何,则这是普通解析几何的更直接的推广,后者本身只不过是代数几何的一部分罢了.代数几何可以认为是数学的这样一个部门,它研究在笛卡儿坐标里由代数方程表示的曲线、曲面和超曲面,这些方程不仅是一次的和二次的,而且还有高次的.原来,在这些研究中最好不仅考虑实数坐标,而且也考虑复

数坐标，即考虑在所谓复空间里的东西．在这个领域里的最重要结果早在上一世纪就由黎曼所得出了．作为关于高阶曲线的定理的光辉的例子，我们提出彼得洛夫斯基的因其普遍性而驰名的结果，他的结果是关于 n 阶曲线所能分成的卵状线个数的．彼得洛夫斯基证明了，如果 p 是完全不在别的卵状线内部和在偶数个卵状线内部的卵状线的个数，而 m 是在奇数个卵状线内部的卵状线的个数，而且如果我们考虑完全由彼此不相交的卵状线组成的曲线(图 92)，则

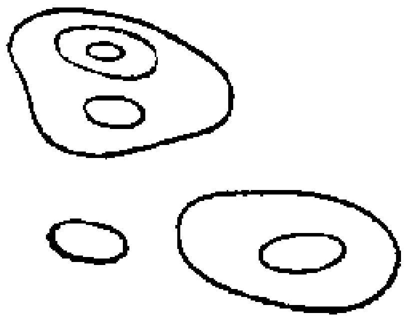


图 92

$$p-m\leqslant\frac{3n^2-6n}{8}+1,$$

这里 n 是曲线的阶数，即是表示它的方程的次数．

上述结果之所以重要，是因为直到今天对于高阶曲线的一般形状几乎是什么也不知道．这恐怕是在解析几何学中最晚得到的重要普遍定理之一．

文　献

通 俗 小 册 子

А. И. Маркушевич, Замечательные кривыт. Гостехиздат, 1952.

А. С. Смогоржевский, Метод координат. Гостехиздат, 1952.

高等技术学校课本

Н. В. 叶菲莫夫，解析几何简明教程，高等教育出版社，1957 年版。

П. И. 普里瓦洛夫，解析几何学，高等教育出版社，1956 年版。

大 学 课 本

Б. Н. 狄隆涅和拉伊可夫合著，解析几何学，第一卷，第一、二分册，高等教育出版社，1956 年版；第二卷，高等教育出版社，1957 年版。

Н. И. 穆斯海里什维里，解析几何教程，高等教育出版社，1954 年版。

裘光明　译

秦元勋　校

第四章 代 数
(代数方程的理论)

§1. 绪 论

什么是代数以及它的特征怎样，由于它的初等性，这个问题已被大家所熟知，并且在中学课程里已给出了代数的基本知识. 代数首先是被它自己的方法刻划出其特点，这些方法就是使用字母及字母的表示式，以及对于它们按照确定的法则进行变换. 在初等代数里，字母代表通常的数，所以字母表示式的变换法则是以数的运算法则为依据的[例如，两数的和与相加的顺序无关，在代数里则写成 $a+b=b+a$；用一数乘二数的和与用此数分别乘二数然后将所得到的两个乘积相加是一样的，即 $(a+b)c=ac+bc$，等等].

我们考察任一代数定理的证明，很容易确信它仅与作用于字母的那些运算法则有关，而与字母所代表的数无关.

代数方法，也就是字母计算的方法贯穿在全部的数学中. 在解决任何数学问题时，其中通常有一部分不是别的，而是一些不同复杂程度的代数计算. 此外，在数学中还使用了各种不同的字母计算法，但此时字母已不一定代表数，而是任何其他对象，因此在这种情况下，字母的运算法则可能是不同于初等的代数，例如，在几何、力学和物理中是使用向量的. 大家都知道向量是有运算法则的，这些法则部分地与数的运算法则相同，但是主要的部分却是不同的.

在近代数学及其应用方面，代数方法的意义在最近几十年来是强有力地增加了.

第一，技术需要的增加要求对一些数学分析的困难问题求出

数字的解答，而这件事通常仅在将这些问题代数化以后才有可能；这样也就构成了新的代数问题，有时甚至是困难的问题.

第二，一些分析方面的问题，在开始应用代数的方法以后，变成明显的和容易理解的，这些方法是建立在一次方程组理论的深刻推广（对于无穷多的情形）的基础上的.

最后，代数的高深部分在近代物理中找到了应用，即量子力学的基本概念是借助于复杂的并且不初等的代数对象表达的.

代数学史的基本特征如下：

首先必须指出的是：关于什么是代数以及代数的基本问题是什么这两个问题的观念有两次改变，一次是上世纪的前半期，而另一次是在本世纪初．在三个不同时期内，人们将三个很不相同的东西理解为代数学，代数学的这些历史不同于解析几何、微分学及积分学这三种著名的计算学科的历史，后者是由它的创始人——费尔马、笛卡儿、牛顿、莱布尼茨及其他学者所亲手奠立的，在进一步的蓬勃发展过程中，甚至有时是用大量的新篇章来补充的时候，它们本来面目在原则上却只有较少的改变.

在古代，为了解决某些种类的数学问题而找到的任何法则都是用语句把它记下来，因为那时字母表示法还没有发明．“代数”这个字本身是由九世纪的花剌子模学者的最重要的著作的名称产生的，这个学者叫做穆罕默德·阿里·花剌子模（参看第一章），在他的著作里产生了第一个解一次及二次方程的一般性法则．然而字母表示法的引进通常是和维耶特的名字相联系的，他不仅用字母表示未知数，并且开始用字母表示给定的量，笛卡儿对于字母表示法的发展也做了不少工作，于是通常的数也可用字母表示．从这个时候开始，实际上把代数看成是关于字母计算，关于由字母构成的公式的变换以及关于代数方程等等的科学，它与算术的不同在于算术永远是对具体数字的运算．仅仅从这以后，甚至很复杂的数学想法都易于加以观察和了解，因为在审视字母的式子以后，在大多数的情况下都能够从它上面看出问题的结构及规律，并且能够容易地加以变换，在那时，整个数学，无论是几何学还是无穷小

分析都被叫做代数学，这就是关于代数学的第一个观点，也就是所谓维耶特的观点，它特别明显地表现在有名的«代数学引论»一书中，这本书是俄罗斯科学院的著名的欧拉院士在十七世纪六十年代写成的，它距离现在已有200年了。

欧拉把代数定义成各种量的计算的理论 他的书的第一部分包含有关于有理整数，通常的分数，二次及三次方根的计算的理论，对数．级数(прогрессия)的理论，多项式的计算理论，牛顿二项式的理论及其应用；在第二部分里包含有一次方程及一次方程组的理论，二次方程的理论以及用根号解三次与四次方程的理论，并且那里面还有各种整数不定方程的解法这类广泛的题材，例如，证明了费尔马方程 $x^3+y^3=z^3$ 不可能有整数解 x, y, z.

在十八世纪末年及十九世纪初年，代数学中的问题之一，即代数方程的解法问题渐渐地被人认为是中心问题．这个问题的根本困难在于一个未知数的 n 次代数方程

$$x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots+a_{n-1}x+a_n=0$$

的解法，这件事的发生是由于这个问题对于整个数学的重要性以及它的应用，同时也由于大多数与它相联系的理论的证明的深刻性和困难性.

一般性的公式

$$x=-\frac{p}{2}\pm\sqrt{\frac{p^2}{4}-q}$$

是人人都知道的，任意的二次方程就是借助于它而解决的，十六世纪的意大利代数学家求出了解三次及四次方程的一般法则，这个法则与上述公式类似，但是却复杂得多，然而对于更高次方程在这件事上的进一步研究却遇到了不可克服的困难．十六、十七、十八以及十九世纪初年的最伟大的数学家们(塔尔塔里雅、卡尔丹、笛卡儿、牛顿、欧拉、达朗贝尔、奇尔恩豪斯、裴蜀、拉格朗日、高斯、阿贝尔、伽罗华、罗巴切夫斯基、斯图谟等等)创造了与这个问题有关的大规模的复杂理论.在十九世纪中叶，谢尔的两卷代数问世了，在这部书里面代数已经定义成为代数方程理论了(这在当时是一个

创举，因为在这部书中第一次叙述了代数方程理论的顶峰——伽罗华理论)．这就是关于什么是代数的第二个观念．

在上一世纪的后半期，由伽罗华的有关代数方程理论的想法出发，群论[1]和代数数论(俄罗斯数学家佐洛塔辽夫在它们的创立中起了巨大的作用)深刻地发展了．

在第二个时期内，作为与代数方程解法问题相关联，同时也与解析几何中所研究的高次代数流形理论相关联的内容，行列式与矩阵的理论，二次型及线性变换的代数理论，特别是不变量的理论等代数工具在各种目的下发展起来了．在几乎整个十九世纪的后半期内，不变量理论是代数研究的中心之一，在这个时期内群论及不变量论的发展也对几何的发展起了重大的影响[2]．

从根本上说，关于什么是代数的第三个观点是与下述的情况相关联的．那就是在上一世纪的后半期内，开始在力学、物理学以及数学本身里越来越频繁地研究到的一些对象，对于这些对象我们很自然地就要考虑它们的加法及减法运算，有时也考虑乘法和除法，然而这些运算满足一些不同于有理数的其他规律．

我们曾经谈到过向量，至于具有其他运算规律的各种形式的量，我们只能在此说出名字，即矩阵、张量、旋量、超复数等等，所有这些量都是用字母表示的，但是不同形式的量的运算规律也有所不同．如果对于某种对象(用字母表示)的某个集合，给出了一些运算以及这些运算所满足的规律(法则)，那末我们就说给出了某一个代数系统．第三个关于什么是代数的观念是说：代数的目的是研究各种代数系统，这就是所谓的公理化的或抽象的代数，它的抽象是由于对所考虑的代数系统是用字母表示，重要之点仅仅是在所考虑的系统里运算满足什么样的公理(规律)．其所以称它为公理化的，是由于使它只服从作为它的基础的那些公理．于是这就回到了第一个(维耶特的)关于什么是代数的观点，即代数是字母计算学的观点，但是已经上升到更高级的形态了．这里对于字

1) 参看第三卷第二十章．

2) 参看第三卷第十七章．

母的了解是相同的，重要的事情仅仅是作用于字母的运算所满足的那些规律，并且有趣的是这样的代数系统无论就数学本身或是在它的应用里都具有巨大的意义.

前一时期所累积的大量的代数内容被认为是创造近代抽象代数的事实基础.

二十世纪三十年代里，范·德·瓦尔登的著名的教科书《近世代数》出版了．它在阐述关于什么是代数的第三个观念上起了巨大的作用．库洛什在同一方面也写了一本代数教科书.

象我们在上面已经谈到的代数在近代物理学(泛函分析及量子力学)上的应用一样，它在本世纪里对于几何(拓扑学及李群的理论)也有深刻的应用.

最近几年来，提出了应用各种数学计算机，特别是快速电子计算机来使代数计算机械化的特别重要的课题，与这些数学机器相关联的各种课题提出了代数学的新的特殊的问题.

在本书中，除了这一章外还有两章是谈论代数的，即线性代数(第三卷第十六章)，群及其他代数系统(第三卷第二十章).

§2. 方程的代数解

形如

$$x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots+a_{n-1}x+a_n=0$$

的方程叫做一个未知数的 n 次代数方程，式中 $a_1,\ a_2,\ \cdots,\ a_n$ 是一些给定的系数[1].

一次及二次方程　如果方程是一次的，那末它的形状是

$$x+a=0,$$

而由此求解即得

$$x=-a.$$

二次方程

$$x^2+px+q=0$$

在远古已经解出，它的解法简单，即将 q 移至右边反号，并且在两

1) 我们假定方程的所有的项都已移至左边，并且用未知数的最高幂的系数除过.

边都加上 $\frac{p^2}{4}$ 就得到

$$x^2+px+\frac{p^2}{4}=\frac{p^2}{4}-q.$$

但

$$x^2+px+\frac{p^2}{4}=\left(x+\frac{p}{2}\right)^2,$$

因此

$$x+\frac{p}{2}=\pm\sqrt{\frac{p^2}{4}-q},$$

由此就得出熟知的二次方程解的公式

$$x=-\frac{p}{2}\pm\sqrt{\frac{p^2}{4}-q}.$$

三次方程 高于二次的方程则是另外一回事了. 一般的三次方程已往需要远非显然的想法，这就使古代数学家的努力都归于无效. 直到十六世纪初的意大利文艺复兴时代，这个问题才被意大利的数学家斯齐波·德尔·菲洛所解决. 按照当时的风气，菲洛的发明没有发表，但却向他的一个学生讲过. 菲洛死后，这个学生用这件事向当时意大利的最大的数学家之一的塔尔塔里雅挑战，要他解出一系列的三次方程. 塔尔塔里雅(1500—1557)起而应战，并且用八天的时间结束了这次竞赛，他得到了解形如 $x^3+px+q=0$ 的任何三次方程的方法.

在两小时之内他解决了对手提出的所有问题. 米兰的数学和物理教授卡尔丹(1501—1576)在得知塔尔塔里雅的发明后，就央求塔尔塔里雅将秘诀告诉他. 塔尔塔里雅终于在要卡尔丹绝对地保守方法的秘密的条件下同意了. 但是卡尔丹背弃了诺言而将塔尔塔里雅的结果发表在他自己的著作《大法》(Ars magna)里.

三次方程的解的公式虽然应该称为塔尔塔里雅公式，但是直到现在为止仍然把它叫做卡尔丹公式.

卡尔丹公式是象下面那样得出的.

首先，解一般三次方程

$$y^3+ay^2+by+c=0 \tag{1}$$

的问题容易归结为解不含未知数的平方项的三次方程

$$x^3+px+q=0 \tag{2}$$

的问题．为此只须令 $y=x-\frac{a}{3}$．事实上，将此式代入方程(1)并去掉括号即得

$$\left(x-\frac{a}{3}\right)^3+a\left(x-\frac{a}{3}\right)^2+b\left(x-\frac{a}{3}\right)+c$$
$$=x^3-3x^2\frac{a}{3}+\cdots+ax^2+\cdots,$$

式中"…"表示 x 的一次及零次各项．由此可见含 x^2 的项是相互抵消了．

现在设给出的方程是

$$x^3+px+q=0.$$

令 $x=u+v$，即引进 u 及 v 后而使一个未知数的问题变成两个未知数的问题．于是就有

$$(u+v)^3+p(u+v)+q=0,$$

或
$$u^3+v^3+q+(3uv+p)(u+v)=0.$$

无论两数和 $u+v$ 是怎样的，我们永远可以要求它们的积等于一个预先给定的值．因为如果给定了 $u+v=A$，而我们要求 $uv=B$，那末由 $v=A-u$ 知要求

$$u(A-u)=B,$$

这只要 u 是二次方程

$$u^2-Au+B=0$$

的根就行了．但由已获得的公式知，每一个二次方程有实数根或复数根．在现今的情形下，$u+v$ 等于我们的三次方程的所求根 x，我们要求

$$uv=-\frac{p}{3},$$

即要求 $3uv+p=0$．对于这样选择的 u 及 v，就得到

$$\begin{aligned}u^3+v^3+q&=0,\\ 3uv+p&=0.\end{aligned}\tag{3}$$

因此我们如果能找到适合这组方程的 u 及 v，那末数 $x=u+v$ 便是原方程的根了．

由方程组容易构成以 u^3 及 v^3 为根的二次方程．事实上，它

给出了

$$u^3+v^3=-q,$$

$$u^3v^3=-\frac{p^3}{27},$$

因此利用上面的定理就知道 u^3 及 v^3 是二次方程

$$z^2+qz-\frac{p^3}{27}=0$$

的根. 按照通常的公式解出它来，就得到

$$u^3=-\frac{q}{2}+\sqrt{\frac{q^2}{4}+\frac{p^3}{27}},\quad v^3=-\frac{q}{2}-\sqrt{\frac{q^2}{4}+\frac{p^3}{27}},$$

故 $$x=\sqrt[3]{-\frac{q}{2}+\sqrt{\frac{q^2}{4}+\frac{p^3}{27}}}+\sqrt[3]{-\frac{q}{2}-\sqrt{\frac{q^2}{4}+\frac{p^3}{27}}}.$$

这就是卡尔丹公式.

四次方程 在三次方程被解出后，一般的四次方程很快地就被费拉里(1522—1565)解出. 为了解出三次方程必需事先解出二次辅助方程

$$z^2+qz-\frac{p^3}{27}=0,$$

式中 $z=u^3$ 或 v^3，而四次方程的解出也有赖于事先解出某一个三次辅助方程.

费拉里的方法如下：设给出的一般四次方程是

$$x^4+ax^3+bx^2+cx+d=0.$$

将它写成 $$x^4+ax^3=-bx^2-cx-d$$

的形式，并且在两边都加上 $\frac{a^2x^2}{4}$，于是左边就是一个完全平方式，即

$$\left(x^2+\frac{ax}{2}\right)^2=\left(\frac{a^2}{4}-b\right)x^2-cx-d.$$

再在这个方程的两边加上

$$\left(x^2+\frac{ax}{2}\right)y+\frac{y^2}{4},$$

式中 y 是一个新的变数，下面将对它给出我们所需要的条件. 这样一来，方程的左边仍然是一个完全平方，即

$$\left(x^2+\frac{ax}{2}+\frac{y}{2}\right)^2=\left(\frac{a^2}{4}-b+y\right)x^2$$
$$+\left(\frac{ay}{2}-c\right)x+\left(\frac{y^2}{4}-d\right). \qquad (4)$$

于是我们将问题化为两个未知数的问题了. 等式(4)的右边是以与 y 有关的式子作系数的 x 的二次三项式. 我们要取一适当的 y 使得这个三项式是二项式 $\alpha x+\beta$ 的平方.

要使二次三项式 Ax^2+Bx+C 是二项式 $\alpha x+\beta$ 的平方, 只要

$$B^2-4AC=0$$

就够了, 事实上, 如果 $B^2-4AC=0$, 那末

$$Ax^2+Bx+C=(\sqrt{A}x+\sqrt{C})^2,$$

即
$$Ax^2+Bx+C=(\alpha x+\beta)^2,$$

式中
$$\alpha=\sqrt{A}, \quad \beta=\sqrt{C}.$$

因此如果取 y 的值使它满足条件

$$\left(\frac{ay}{2}-c\right)^2-4\left(\frac{a^2}{4}-b+y\right)\left(\frac{y^2}{4}-d\right)=0,$$

那末方程(4)的右边就是完全平方 $(\alpha x+\beta)^2$. 展开此式, 我们得到 y 的一个三次方程, 即

$$y^3-by^2+(ac-4d)y-[d(a^2-4b)+c^2]=0.$$

解出这个辅助的三次方程后(例如按照卡尔丹公式), 然后由它的根 y_0 求出 α 及 β, 于是就有

$$\left(x^2+\frac{ax}{2}+\frac{y_0}{2}\right)^2=(\alpha x+\beta)^2,$$

由此得到

$$x^2+\frac{ax}{2}+\frac{y_0}{2}=\alpha x+\beta \quad 或 \quad x^2+\frac{ax}{2}+\frac{y_0}{2}=-\alpha x-\beta.$$

由这两个二次方程, 我们就求出所给出的四次方程的四个根.

十六世纪的意大利数学家对三次及四次代数方程的解答就是这样.

意大利数学家的成就发生了很大的影响. 在当时, 新时代的科学还是第一次超过了旧时的成就. 整个中世纪只是处于了解古

代著作的潮流影响之下，而终于在这里解决了古代所不能解决的问题．这是在新的计算学科——解析几何、微分学及积分学被发现的前一百年的时候，这些新的计算学科最终地肯定了新的科学比旧的优越．在这些伟大数学家之后，人们并不是不设法继承意大利人的成功而用根号来类似地解出五次、六次以及更高次的方程．

十七世纪杰出的代数学家奇尔恩豪斯(1651—1708)甚至说他终于找到了一般的解法．他的方法是建筑在将一方程变换为更简单的方程的基础上，但这个变换本身需要解某些辅助方程，经过深入地考察以后，发现奇尔恩豪斯的变换方法实际上只给出了二次、三次及四次方程的解答，而对于五次方程则需要事先解出一个还不知如何去解的六次辅助方程．

分解因式及维耶特公式　如果应用所谓的代数基本定理[1)]（先不管它的证明），即设

$$f(x)=x^n+a_1x^{n-1}+\cdots+a_n$$

是任意给定的 x 的 n 次多项式，它的系数 a_1, a_2, …, a_n 是给定的实数或复数，则方程

$$f(x)=0$$

至少有一实数或复数根，其中复数的计算是按照有理数所适合的规则一样地进行，那末容易证明多项式 $f(x)$ 能够表成一次因式的连乘积的形状，即

$$f(x)=(x-a)(x-b)\cdots(x-l),$$

此处 a, b, …, l 是一些实数或复数，并且只有这一种表示法．

事实上，设 a 是 $f(x)$ 的一根，以 $x-a$ 去除 $f(x)$，由于除式是 1 次的，因此余数就是一个常数 R，即我们有恒等式

$$f(x)=(x-a)f_1(x)+R,$$

式中 $f_1(x)$ 是一个 $n-1$ 次的多项式，而 R 是常数，用数 a 代 x 以后就得到

1) 代数基本定理的证明是困难的，并且是在更晚得多的时候给出的，在这里将以 §3 的整个篇幅来完成它，但它的正确性却在严格证明很久以前就料想到了．

$$f(a)=(a-a)f_1(a)+R=R.$$

因为 a 是 $f(x)$ 的一根，故 $f(a)=0$，所以 $R=0$. 即 a 如果是一个多项式的根，$x-a$ 就能整除此多项式，故

$$f(x)=(x-a)f_1(x).$$

如果代数基本定理是正确的话，那末 $f_1(x)$ 也有一个根 b，类似地我们就得到

$$f_1(x)=(x-b)f_2(x),$$

式中 $f_2(x)$ 是一个 $n-2$ 次的多项式等等. 这个分解式容易证明是唯一的.

每一个 n 次多项式 $f(x)$ 在这种意义下有 n 个并且有 n 个根 $a, b, c, \cdots, l$. 它们可能全不相等，也可能有一些是相同的. 当有一些根是相同的时候，我们就把相互相同的根的个数叫做这个根的重复度.

将式子 $\qquad (x-a)(x-b)(x-c)\cdots(x-l)$

乘开，并比较 x 的同次幂的系数，我们就直接得到

$$-a_1=a+b+c+\cdots+l,$$
$$a_2=ab+ac+\cdots+kl,$$
$$-a_3=abc+abd+\cdots,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\pm a_n=abc\cdots l.$$

这就是维耶特公式.

有关对称多项式的定理 维耶特公式是 n 个字母 $a, b, \cdots, l$ 的多项式，它对这些字母的任何排列不变. 实际上，$a+b+\cdots+k+l=b+a+\cdots+k+l$，等等，一般来说对 n 个字母的任何排列不变的 n 个字母的多项式叫做这 n 个字母的对称多项式. 例如，$5x^2+5y^2-7xy$ 是 x 及 y 的对称多项式，能够证明下面的定理：每一具有任何系数 $A, B, \cdots$ 的 n 个字母的整对称多项式能够表成这些字母的维耶特多项式及系数 $A, B\cdots$ 的有理整式(即应用加、减、乘诸运算而得到的式子). 如果 $a, b, \cdots, l$ 是 n 次方程 $x^n+a_1x^{n-1}+\cdots+a_n=0$ 的根，那末具任何系数 $A, B, \cdots$ 的 $a, b, \cdots, l$

的每一对称多项式就是方程的系数 $a_1, a_2, \cdots, a_n$ 及系数 $A, B, \cdots$ 的有理整式．这就是对称多项式的基本定理．

拉格朗日的工作 著名的法兰西数学家拉格朗日在1770—1771年所发表的长文(有200多页)“关于代数方程解法的思考”中，讨论了在他以前为人们所熟知的解二次、三次及四次方程的一切解法，他并且指出这些成功了的解法所根据的情况对五次及更高次方程是不可能发生的．从菲洛所处的时代到拉格朗日的这篇文章出版时，中间经过了两个半世纪，在这样长的时间里任何人都没有怀疑过用根号解五次及更高次方程的可能性，也就是说，大家认为可以找到一个表示这些方程的根的公式，而这个公式象古代的解二次方程及十六世纪意大利人解三次及四次方程一样，只是对这些方程的系数作加、减、乘、除及求正整数次根诸运算就可得到．仅仅以为是人们没有能有成效地找到正确的然而看来是很诡秘的道路去得出解法．

拉格朗日在他自己的回忆录(全集第三卷第305页)中曾说“用根号解四次以上的方程的问题是不可能解决的问题之一，虽然关于解法的不可能性什么事情也没有证明”，在第307页他补充说“由我们的研讨可以看出用我们所考虑的方法给出五次方程的完全解法是很值得怀疑的”．

在拉格朗日的研究中，他引进了式子，

$$a+\varepsilon b+\varepsilon^2 c+\cdots+\varepsilon^{n-1} l,$$

式中 $a, b, c, \cdots, l$ 是方程的根，ε 是1的任一 n 次根[1]，并且确定了正是这些式子紧密地联系着用根号解方程，现在我们将这些式子叫做“拉格朗日预解式”．

---

1) 即它们的 n 次方等于1的那些数．例如，1的立方根是

$$1,\ -\frac{1}{2}+\frac{\sqrt{3}}{2}i,\ -\frac{1}{2}-\frac{\sqrt{3}}{2}i,$$

式中 $i=\sqrt{-1}$ (参看第293页)．事实上，

$$\left(-\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)^3=-\frac{1}{8}-\frac{3}{8}\sqrt{3}\,i+\frac{9}{8}+\frac{3\sqrt{3}}{8}i=1,$$

类似地有 $\left(-\frac{1}{2}-\frac{\sqrt{3}}{2}i\right)^3=1$．

此外，拉格朗日觉察到方程的根的排列理论比方程用根号解的理论有更大的意义，他甚至表达出排列的理论是"整个问题的真正哲学"的看法，正如后来伽罗华的研究所指出的那样，他是完全正确的.

拉格朗日导出二次、三次及四次方程的解并不象意大利人那样，对每种情况都有它某种固有的复杂性并且好象是偶然地找到的一些变换，相反地他是十分严格地并且是从一个一般的想法借助于对称多项式的理论、置换的理论及预解式理论的统一的方法导出的.

例如，让我们来考察一下拉格朗日对一般四次方程

$$x^4+mx^3+nx^2+px+q=0$$

的解法.

设这个方程的根是 a, b, c, d. 我们考虑预解式

$$a+b-c-d,$$

即

$$a+\varepsilon c+\varepsilon^2 b+\varepsilon^3 d,$$

式中 $\varepsilon=-1$. 用 $1\cdot2\cdot3\cdot4=24$ 种不同的方法作出 a, b, c, d 的一切排列，我们就得到六个不同的式子

$$\begin{aligned}&a+b-c-d,\\&a+c-b-d,\\&a+d-c-b,\\&c+d-a-b,\\&b+d-a-c,\\&b+c-a-d.\end{aligned}\tag{5}$$

以这六个式子为根的六次方程的系数是对 a, b, c, d 的所有24个排列不变的，因为这24个排列中的任何一个仅仅只能将这些式子重新排列一下，而所考虑的六次方程的系数并不依赖于它的根的先后次序. 这样，这些系数就是 a, b, c, d 的对称多项式了，由于对称多项式的基本定理，这些系数可以用方程的系数 m, n, p, q 有整理地表示出来. 此外因为(5)式是成对地具有相反符号的，所以这个六次方程仅仅包含偶数次项. 事实上，如果(5)式

用 $\alpha, \beta, \gamma, -\alpha, -\beta, -\gamma$ 来记 那末所考虑的六次方程的左边就等于

$$(y-\alpha)(y+\alpha)(y-\beta)(y+\beta)(y-\gamma)(y+\gamma)$$
$$=(y^2-\alpha^2)(y^2-\beta^2)(y^2-\gamma^2).$$

由直接的计算就知道这个六次方程是

$$y^6-(3m^2-8n)y^4+3(m^4-16m^2n-16n^2+16mp-64q)y^2$$
$$-(m^3-4m+18p)^2=0.$$

令 $y^2=t$, 就得到 t 的三次方程, 假设 t', t'', t''' 是它的根, 那末

$$a+b-c-d=\sqrt{t'},$$
$$a+c-b-d=\sqrt{t''},$$
$$a+d-b-c=\sqrt{t'''}.$$

此外还有 $$a+b+c+d=-m.$$

用 1, 1, 1, 1, 或 1, -1, -1, 1, 或 -1, 1, -1, 1, 或 -1, -1, 1, 1 依次乘以上四个方程然后逐项相加就得到

$$a=\frac{1}{4}(-m+\sqrt{t'}+\sqrt{t''}+\sqrt{t'''}),$$

$$b=\frac{1}{4}(-m+\sqrt{t'}-\sqrt{t''}-\sqrt{t'''}),$$

$$c=\frac{1}{4}(-m-\sqrt{t'}+\sqrt{t''}-\sqrt{t'''}),$$

$$d=\frac{1}{4}(-m-\sqrt{t'}-\sqrt{t''}+\sqrt{t'''}).$$

这样一来,四次方程的解法就归结为三次方程的解法,类似地可以去解二次及三次方程.

拉格朗日在代数方程的理论方面的成就很多, 然而经过他顽强的努力之后,用根号解高于四次的方程的问题仍然悬而未决.这个几乎费了三个世纪的劳动的问题正如拉格朗日所表述的一样"它好象是在向人类的智慧挑战".

阿贝尔的发现 在 1824 年一个年青的有天才的挪威人阿贝尔(1802—1829)的著作出版时,引起了所有的数学家的惊奇,在这个著作中证明了这样一件事:如果方程的次数 $n\geqslant 5$, 并且系数 a_1,

$a_2, \cdots, a_n$ 看成是字母，那末任何一个由这些系数组成的根式不可能是方程的根．原来一切国家的最伟大的数学家三个世纪以来用根号去解五次或更高次方程之所以不能获得成就，只因为这个问题根本就没有解．

对于二次方程这样的公式是大家都知道的；对于三次及四次方程正如我们所见到的，也有类似的公式，然而对于五次及更高次方程，就没有任何这样的公式存在．

阿贝尔的证明很难，我们将不再在此引用．

伽罗华理论 然而这并不是问题的全部，代数方程的理论的最美妙之处仍然留在前面，问题在于有多少种特殊形式的方程能够用根号解，而这些方程又恰恰在多方面的应用中是重要的．例如，二项方程 $x^n=A$ 便是这样的方程．阿贝尔找到很广泛的另一类这样的方程，就是所谓的循环方程及更一般的"阿贝尔"方程．由于用圆规直尺作正多角形的问题，高斯详尽地考察了所谓的分圆方程，也就是形如

$$x^{p-1}+x^{p-2}+\cdots+x+1=0$$

的方程，其中 p 是素数，证明了它们总能归为解一串较低次的方程，并且他找到了这种方程能用二次根号解出的充分与必要条件（这个条件的必要性的证明只是到伽罗华时才有了严格的基础）．

总之，在阿贝尔的工作之后的情况就是这样：虽然阿贝尔证明了高于四次的一般方程是不能用根号解出，但是有多少种不同类型的特殊的任何次方程仍然是可以用根号解出呢？由于这个发现，关于用根号解方程的全部问题是在新的基础上提出来了，应该找出一切能用根号解出的那些方程，换句话说，就是找出方程能用根号解出的充分与必要的条件，这个问题是由天才的法兰西数学家埃瓦里斯特·伽罗华解决的，而问题的答案在某种意义下给出了全部问题的彻底的阐明．

伽罗华(1811—1832)在20岁的时候与人决斗而死，在他的生命的最后两年中，因为他热衷于1830年革命时的狂风暴雨式的政治生活，由于他反对路易-菲利普的反动制度的言论而坐牢等等，

他不可能对数学这一学科花费很多的时间，然而伽罗华在自己的生命中却在数学的一些不同的部门中给出了远远超过他的时代的发明，特别是给出了代数方程论中最著名的结果．在伽罗华死后所留下的并由刘徵尔在1846年首次发表的不算长的手稿“关于方程用根号解的条件的记录”中，从很简单的但是很深刻的想法出发，终于解开了环绕着用根号解出方程的困难的症结，而这个困难却是伟大的数学家们所毫无成效地奋斗过的．伽罗华的成就在于他在方程论中第一次引进了非常重要的新的一般概念，这些概念后来在整个数学中起着重要的作用．

我们现在就一种情形来考察伽罗华理论，即考察系数 $a_1, a_2, \cdots, a_n$ 是有理数的 n 次方程

$$x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n=0. \tag{6}$$

这个情形是特别有趣的，并且实际上已经包含了伽罗华一般理论的全部困难．此外，我们假设所考虑的方程的根 $a, b, c, \cdots$ 是不同的．

与拉格朗日所考虑的很类似，伽罗华从考虑关于 $a, b, c, \cdots$ 的某种一次式

$$V=Aa+Bb+Cc+\cdots$$

开始，但他不必假设这个式子的系数 $A, B, C, \cdots$ 是单位根，而取 $A, B, C, \cdots$ 为使得在 V 中对根 $a, b, c, \cdots$ 作一切可能的 $n!=1\cdot2\cdot3\cdots n$ 个排列所得到的 $n!$ 个值 $V, V', V'', \cdots, V^{(n!-1)}$．不同的某些有理整数，这一点是永远可以办到的．其次，伽罗华作出以 $V, V', V'', \cdots, V^{(n!-1)}$ 为根的方程．借助于关于对称多项式的定理，不难证明这个 $n!$ 次方程 $\Phi(x)=0$ 的系数是有理数．

到此为止，都很类似于拉格朗日所曾经做过的．

进一步伽罗华引进了第一个新的重要的概念，即在给定的数域上不可约多项式的概念．如果给了 x 的某一个多项式，它的系数比方说是有理数，那末如果这个多项式能表成更低次多项式的乘积的形式，这个多项式就叫做在有理数域上可约．多项式 x^3-x^2-4x-6 在有理数域上便是可约的，因为它等于 $(x^2+2x+2)(x-3)$，

但是能够证明多项式 x^3+3x^2+3x-5 在有理数域上不可约.

虽然需要很长的计算，但是将任一有理系数多项式分解成在有理数域不可约多项式的乘积的方法是存在的.

伽罗华假定我们所得到的多项式 $\Phi(x)$ 已经分解成有理数域上不可约多项式之乘积.

设 $F(x)$ 是这样的不可约因子中任意给定的一个，并且设它的次数是 m.

此时多项式 $F(x)$ 就是 $n!$ 次多项式 $\Phi(x)$ 的 $n!$ 个一次因子 $x-V,\ x-V',\ \cdots,\ x-V^{(n!-1)}$ 中的 m 个的乘积，假设这 m 个因子是 $x-V,\ x-V',\ \cdots,\ x-V^{(m-1)}$. 将给定的 n 次方程(6)的根 $a,\ b,\ \cdots,\ l$ 用数码 1, 2, …, n 标出，那末 $V,\ V',\ \cdots,\ V^{(n!-1)}$ 就由根的号码 1, 2, …, n 的所有可能的 $n!$ 排列所引出，而 $V,\ V'\cdots,\ V^{(m-1)}$ 仅仅是它们中间的 m 个. 号码 1, 2, …, n 的这 m 个排列的全体 G 就叫做给定的方程(6)的伽罗华群[1].

伽罗华还引进了一些新的概念，并且通过它们得到了方程(6)可用根号解出的充分与必要条件是号码 1, 2, …, n 的排列群 G 满足某些确定的条件. 这些论证虽然简单，但的确是美妙的论证.

由此可见拉格朗日指出整个问题是以排列理论为基础的预见是正确的.

特别地，阿贝尔的关于一般五次方程不能用根号解出的定理现在可以象下面那样证明：能够证明有任意多个具整系数的五次方程使得它所对应的 120 次的方程 $\Phi(x)$ 是不可约的，这就是说它们的伽罗华群是号码 1, 2, 3, 4, 5 的所有的 $5!=120$ 个排列的群. 但是能证明这个群不满足伽罗华的判别条件，所以一般五次方程不能用根号解出.

例如，若 a 是正整数，能够证明方程 $x^5+x-a=0$ 大部分不能用根号解出，例如当 $a=3,\ 4,\ 5,\ 7,\ 8,\ 9,\ 10,\ 11,\ \cdots$ 时，它们不能用根号解出.

伽罗华理论对用圆规直尺作几何图形的问题的应用 下面是

1) 关于伽罗华群将在第三卷第二十章 § 5 中更详细地谈到.

伽罗华理论卓越的应用之一．很多平面几何的问题能用圆规直尺作图解出，而另外一些就不行．例如能用圆规直尺作正三角形，正方形，正五角形，正六角形，正八角形，正十角形等等，但是不能作正七角形，正九角形，正十一角形等等．那末什么样的问题能用圆规及直尺解出，而什么样的问题又不能呢？这是一个在伽罗华以前没解决的问题，而由伽罗华理论就得到它的下列的解答．

两条直线的方程，一条直线的方程与圆周的方程或者两个圆周的方程的公共解归结为一次或二次方程的解．对于直线与圆周来说，这一点是显然的；而对于两个圆周$(x-a_1)^2+(y-b_1)^2=r_1$及$(x-a_2)^2+(y-b_2)^2=r_2^2$的情形，如果从一个方程减去另一方程，那末$x^2$及$y^2$消去了，而得到一个一次方程，将它与一个圆周联合求解，同样得到一个二次方程．所以每一个能用圆规及直尺解出的问题归结到一次或二次方程，因此整个问题就归结为一个未知数的代数方程，而它的解可由一串平方根导出．反之，如果一个几何问题归结为这样的代数方程，那末它就能用圆规及直尺解出，因为大家所熟知的，平方根是能用圆规直尺作出来的．

如果给了任意一个几何问题，就需要作出它所归结到的那个代数方程．如果这样的代数方程不能造出来，那末问题不能用圆规及直尺解出就是很明显的事．如果这样的代数方程存在，那末就需要分出与问题解答相关联的不可约因子，并且查明这个不可约方程是否能用平方根号解出．如伽罗华理论所表明的，这件事成立的充分与必要条件是组成方程的伽罗华群的排列的个数m是2的方幂．

借助这个判别条件，证明了高斯所表达过的下列定理：具有素数p个边的正多角形可以用圆规直尺作出的充分与必要条件是p具有$2^{2^n}+1$的形式，即对$p=3, 5, 17, 257$等情形能够用圆规直尺作出，而对$p=7, 11, 13, 19, 23, 29, 31$等等则不可能用圆规直尺作出．高斯仅仅证明了第一个论断．

用同样的方法还能证明：用圆规及直尺不能三等分一任何角；不能解决立方倍积的问题，即由一给定的立方体的棱出发找出另

一立方体的棱使所找出的立方体的体积两倍于给定的立方体的体积；等等.

化圆为方的不可能性，即用圆规及直尺不可能由一给定的圆的半径作出一与此圆等面积的正方形的边则是另外证明的．就是说能够证明这样的正方形的边不可能用任何一个代数方程与圆的半径联系起来，即如通常所说的，它对圆的半径是超越的，因此更谈不到用半径的一串平方根号来表出．这件事的证明是困难的，并且它不是从伽罗华理论导出的.

与伽罗华理论有关的两个基本问题 在伽罗华理论中留下了两个基本的进一步的问题，虽然很多杰出的数学家对它们进行过不倦的努力，但是到现在为止还不能在一般的形式下解决.

第一个问题就是所谓的希尔伯特-切波塔雷夫预解式（不要与拉格朗日预解式混同起来）问题，它是用根号解方程的直接推广．事实上用根号解出方程的说法与方程的解答可以归结为解一串二项方程的说法是相同的，因为根号 $\sqrt[n]{A}$ 与二项方程 x^n-A 的根是一回事．很可能的是：即令方程不能归结为象二项方程这样简单的方程，但仍然能归结为另一串很简单的方程．早在十八世纪末人们已经证明了一般的五次方程能够归结为一串二项方程还外加一个形如 $x^5+x+A=0$ 的方程，虽然它不是二项方程，但它仅有一个参数 A，也就是说它与二项方程同样只有一个参数.

后来又证明了六次方程不能归结为一串具有一个参数的方程．这就需要对于每一次数的方程来考虑它能归结为一串什么样的最简单的方程，即具有最少参数的方程.

如果给定的方程归结为一串具有确定形式的具有一个参数的方程，那末对每一个这样具有一个参数的方程能够按照它的参数值算出它的根的表．于是解出给定方程就归结为应用一串这样的表.

第二个问题是伽罗华理论所要求的更深刻的问题．伽罗华证明方程的解的性质依赖于它的群．反之，是否每一排列群一定是某一方程的伽罗华群，并且以它为群的一切方程是怎么样的.

关于第一个问题，象大家知道的，即令是象克莱茵和希尔伯特这样杰出的数学家对它作了顽强的努力后，也只得到一些个别的结果；第一个一般性定理是著名的苏联代数学家切波塔雷夫所证明的.

在最近几年，苏联数学家沙法列维奇对第二个问题的所谓可解群的情形，即满足伽罗华判别条件的情形，作了肯定的回答.

§3. 代数基本定理

在前一节里，我们考察了曾经经历了三个世纪的想用根号解出 n 次方程的企图. 问题的确是很困难和很深刻的；而且引导到建立了不仅是对代数而且对整个数学都很重要的新的想法. 至于谈到方程的实际解法，那末可以对所有这些巨大的工作作出如下的总结. 已经查明用根号解方程的方法对一切代数方程来讲是远远不够的；并且除了二次方程以外，即令是能用根号解的情形也由于它的复杂性而对应用来说是很少有用的.

由于这个原因，数学家对代数方程理论早就着手在其他三个方向进行工作，即：1)关于根的存在的问题；2)不解出方程而按照它的系数去探索它的根的一些性质，例如它是否有实根的问题，有多少个的问题；3) 关于方程的根的近似计算的问题.

首先需要证明的是：每一实系数或复系数 n 次代数方程永远至少有一实根或复根[1].

这个定理是整个数学中最重要的定理之一，但是很久没有严格的证明. 由于它的证明的困难及具有基本性质，通常叫做“代数基本定理”. 而就实质来说，证明的方法的大部分与其说是代数的，不如说是无穷小分析的. 它的第一个证明是达朗贝尔给出的. 达朗贝尔的证明在后来发现有一点是不充分的. 这是说达朗贝尔应用了数学分析中一条甚为明显的一般性引理，即在一有限闭区间上定义的连续函数一定在某一点取极小值. 这件事是对的，但

1) 事实上，有非代数的方程存在，例如 $a^x=0$，它既没有实根也没有复根.

是需要证明. 这个命题的严格证明在十八世纪的后半期才得到, 即在达朗贝尔的研究一百年以后才得到.

代数基本定理的第一个严格证明通常认为是高斯给出的; 然而在他的证明的某些地方, 为了完全严格所需要作出的补充并不少于达朗贝尔的证明. 现在已经有了这个定理的一系列的不同的完全严格的证明.

在本节里我们考察代数基本定理的基于达朗贝尔引理的证明, 并且引进上面提到的数学分析上的引理的完全证明.

复数的理论 在考察代数基本定理以前, 首先必需回忆我们在中学里已经学过的复数理论. 引导人们去建立复数的困难已经在解二次方程时首次遇到了. 即在二次方程解的公式中, 平方根号下的数 $\frac{p^2}{4}-q$ 是负数时应该怎么办呢? 任何一个实数, 不管是正数还是负数都不可能是负数的平方根, 因为任何实数的平方是正数或零.

经过了一世纪以上的怀疑后, 数学家们得出了必须引进具有下列运算法则的新型的数的结论, 这种数就是复数.

约定所引进的新型的数 $i=\sqrt{-1}$ 使得 $i^2=-1$ 后, 考察形如 $a+bi$ 的数, 此处 a 及 b 是通常的实数. 数 $a+bi$ 叫做复数. 两个这样的数 $a+bi$ 及 $c+di$ 当 $a=c$, $b=d$ 时算作相等. 数 $(a+c)+(b+d)i$ 算作是这样的两数之和, 而 $(a-c)+(b-d)i$ 是差. 对于乘法则将此两数看作是两个二项式相乘, 然后应用 $i^2=-1$, 即

$$\begin{aligned}(a+bi)(c+di)\\&=ac+bci+adi+bdi^2\\&=(ac-bd)+(bc+ad)i.\end{aligned}$$

如果把 a 与 b 认为是点的直交坐标, 而这个点与复数 $a+bi$ 相对应, 那末我们所指出的复数的加法与减法对应于自原点到具有坐标 (a, b), (c, d) 的点的向量(有向线

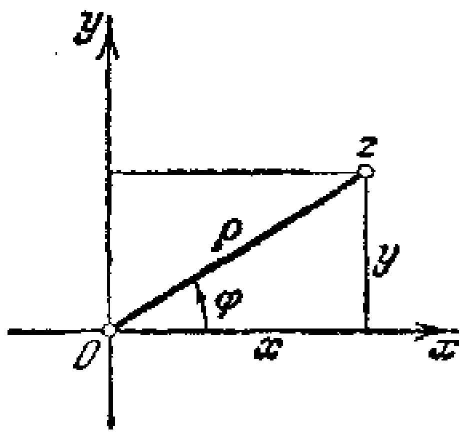


图 1

段)的加法，因为向量的加法是由它们的对应的坐标的加法作成的．至于上述的乘法在这个所谓的复数平面究竟具有什么样的意义，如果我们不考察对应于复数 $z=x+iy$ 的点的笛卡儿坐标 x 和 y，而考察它的极坐标 ρ 与 φ 时，就比较容易看出了，即考察自原点到点(x, y)的向量的长(这个长叫做复数 $z=x+iy$ 的模)及这个向量与 Ox 轴所成的角 φ(这个角叫做复数 $z=x+iy$ 的辐角)(图1)．此时 $x=\rho\cos\varphi$, $y=\rho\sin\varphi$, 因此复数本身就写成

$$x+iy=\rho(\cos\varphi+i\sin\varphi).$$

如果

$$a+bi=\rho_1(\cos\varphi_1+i\sin\varphi_1),\quad c+di=\rho_2(\cos\varphi_2+i\sin\varphi_2),$$

那末

$$ac-bd=\rho_1\rho_2(\cos\varphi_1\cos\varphi_2-\sin\varphi_1\sin\varphi_2)=\rho_1\rho_2\cos(\varphi_1+\varphi_2),$$

$$bc+ad=\rho_1\rho_2(\sin\varphi_1\cos\varphi_2+\cos\varphi_1\sin\varphi_2)=\rho_1\rho_2\sin(\varphi_1+\varphi_2),$$

由此可见，当两个复数相乘时，它们的模 ρ_1 及 ρ_2 相乘，而辐角 φ_1 及 φ_2 相加．由于除法是乘法的逆运算，所以对于除法是一数的模去除另一数，而辐角相减．即

$$\rho_1(\cos\varphi_1+i\sin\varphi_1)\rho_2(\cos\varphi_2+i\sin\varphi_2)$$
$$=\rho_1\rho_2[\cos(\varphi_1+\varphi_2)+i\sin(\varphi_1+\varphi_2)]$$

及 $$\frac{\rho_1(\cos\varphi_1+i\sin\varphi_1)}{\rho_2(\cos\varphi_2+i\sin\varphi_2)}=\frac{\rho_1}{\rho_2}[\cos(\varphi_1-\varphi_2)+i\sin(\varphi_1-\varphi_2)].$$

因此对于正整数次的 n 次幂，它的模就是原来的模的 n 次幂，而辐角是原来辐角的 n 倍，

$$[\rho(\cos\varphi+i\sin\varphi)]^n=\rho^n(\cos n\varphi+i\sin n\varphi).$$

反之对于开方则有

$$\sqrt[n]{\rho(\cos\varphi+i\sin\varphi)}=\sqrt[n]{\rho}\left(\cos\frac{\varphi}{n}+i\sin\frac{\varphi}{n}\right).$$

然而开方毕竟发生了一种特殊情况．设 n 是正整指数，则

$$\sqrt[n]{\rho(\cos\varphi+i\sin\varphi)}$$

等于数 $$\sqrt[n]{\rho}\left(\cos\frac{\varphi}{n}+i\sin\frac{\varphi}{n}\right),$$

因为这个数的 n 次方给出了被开方的数.

但是这仅仅是根的一个值. 问题在于复数

$$\sqrt[n]{\rho}\left[\cos\left(\frac{\varphi}{n}+\frac{2k\pi}{n}\right)+i\sin\left(\frac{\varphi}{n}+\frac{2k\pi}{n}\right)\right]$$

也是数

$$\rho(\cos\varphi+i\sin\varphi)$$

的 n 次根, 式中 k 是 1, 2, …, $n-1$ 中的任一数. 事实上, 按照乘方规则将这样的数作 n 次方时得到数

$$(\sqrt[n]{\rho})^n\left[\cos n\left(\frac{\varphi}{n}+\frac{2k\pi}{n}\right)+i\sin n\left(\frac{\varphi}{n}+\frac{2k\pi}{n}\right)\right]$$
$$=\rho[\cos(\varphi+2k\pi)+i\sin(\varphi+2k\pi)],$$

并且在余弦和正弦的符号下的 $2k\pi$ 这一项可以不写, 因为它使正弦及余弦都不变. 这样, 这个数的 n 次幂是数

$$\rho(\cos\varphi+i\sin\varphi),$$

即这个数是

$$\sqrt[n]{\rho(\cos\varphi+i\sin\varphi)},$$

容易看出除了这 n 个数(当 $k=0$, 1, 2, …, $n-1$)以外, 其他任何复数不再是

$$\rho(\cos\varphi+i\sin\varphi)$$

的 n 次根. 几何地求 n 次根表述如下. 与数 $\rho(\cos\varphi+i\sin\varphi)$ 的 $\sqrt[n]{\ }$ 的值对应的复平面上的点是一正 n 角形的顶点, 这些顶点在以原点为心以 $\sqrt[n]{\rho}$ 为半径的圆上, 而这个 n 角形的顶点之一有辐角 $\frac{\varphi}{n}$(图 2).

我们给出下面的注意. 若

$$f(z)=z^n+c_1z^{n-1}+\cdots+c_{n-1}z+c_n$$

是具有实或复系数 c_1, c_2, …, c_n 的 z 的多项式, 并让 z 连续地变动, 即点 $z=x+iy$ 在复平面上连续地移动, 此时复数点 $Z=X+iY=f(z)$ 在复平面上也连续地移动. 这件事由下面的论证就看得出是很明显的: 如果以 $z=x+iy$, $c_1=a_1+b_1i$, $c_2=a_2+b_2i$, …, $c_n=a_n+b_ni$ 代入 $f(z)$ 并且进行计算, 那末就得到

$$f(z)=X+iY,$$

此处

$$X=P(x, y),\ Y=Q(x, y)$$

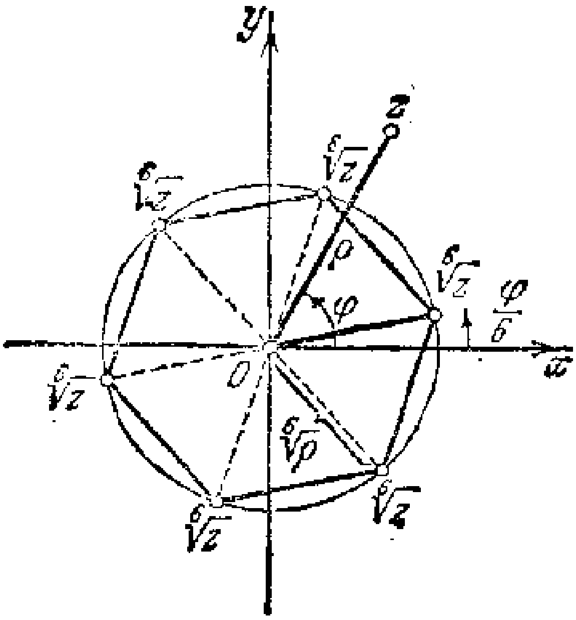


图 2

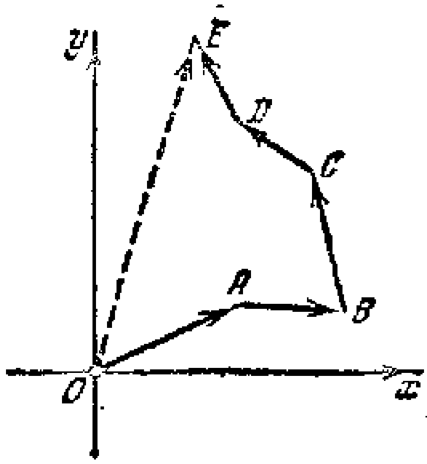


图 3

是 x 及 y 的实系数多项式，而系数是由 a_i 及 b_i 表出的，对于连续的变量 x 与 y，这些多项式也连续地变动。

我们还指出因为模 $\rho=|f(z)|$ 等于 $\sqrt{X^2+Y^2}$，所以当点 z 在复平面上连续地移动时，模 $|f(z)|$ 也连续地变动，换句话说，如果点 z 充分地接近点 a，那末差 $|f(z)|-|f(a)|$ 的绝对值可以小于任何预先给定的正数。

我们还指出若干个复数的和的模永远小于或等于这些数的模的和，这等价于折线 $OABCDE$ 的封闭直线线段 OE（图 3）短于该折线，并且当且仅当折线的所有线段在一条直线上且沿同一方向延伸时相等。

最后，我们记出"复数等于零"或"它的模等于零"是同义的，因为复数的模 ρ 是对应的点到零点的距离。

我们现在应用复数的理论去证明代数基本定理，然而复数理论的意义远远超过代数的范围．象在代数学里一样，在数学的很多其他部门里没有它是不行的．在很多应用中，例如在交流电的理论中，一系列的问题是借助于复数得到最简单的解决．但是复数运用到两个实变数的某些特殊的函数即所谓调和函数的理论上去是最重要的，这件事实际上就是复变函数理论．借助于这些函

数解决了飞机飞行理论，热运动理论，电场理论的重要问题及弹性理论的某些问题. 关于飞机机翼升力的著名定理是由近代空气动力学的创始人茹可夫斯基借助于复变函数论的研究而得到的[1].

现在转到代数基本定理的证明上来.

定理 以任意给定的实数或复数

$$a_0,\ a_1,\ \cdots,\ a_{n-1},\ a_n$$

为系数的多项式

$$f(z)=a_0z^n+a_1z^{n-1}+\cdots+a_{n-1}z+a_n,$$

至少有一实数根或复数根.

我们将假定给定的多项式是 n 次的，即 $a_0\neq 0$.

多项式的模的曲面 我们几何地考察全部问题. 在每一个复数点 z 上安置立坐标(垂线) t，它的长等于多项式 $f(z)$ 在这一点 z 的模 $|f(z)|$. 这些立坐标的端点形成某一曲面 M，我们把它叫做多项式 $f(z)$ 的模的曲面，我们发现这个曲面有下列性质：1)无论在什么地方，它不会降到复平面下面，这是因为任何复数[目前情况是数 $f(z)$] 的模是非负的；2)对于复平面的任何一点 z，这个曲面上有一点且仅有一点位于这个点的上面或者就是这个点本身；即曲面 M 在整个复平面上是由一叶构成的，并且可能某些点与此平面接触；3) 这个曲面在下述意义下连续，当点 z 的位置在复平面上连续变动时，这个曲面的点的立坐标 $t=|f(z)|$ 连续地变动(这一点在前一段已经证明).

代数基本定理的要求就是证明曲面 M 至少在一点接触到复平面，而不是处处都是在复平面上面的某一高度经过.

离原点渐远多项式的模的递增 我们要证明无论给出一个多么大的数 G，总能找出一个 R 使得与位于以原点为圆心以 R 为半径的圆外的点 z 对应的曲面 M 上的点的立坐标 t 大于 G.

事实上，将多项式 $f(z)$ 写成

$$a_0z^n\left[1+\left(\frac{a_1}{a_0z}+\frac{a_2}{a_0z^2}+\cdots+\frac{a_n}{a_0z^n}\right)\right]$$

---

1) 参照第二卷第九章.

的形式. $\left(\frac{a_1}{a_0z}+\frac{a_2}{a_0z^2}+\cdots+\frac{a_n}{a_0z^n}\right)$的模不大于各项的模的和 $\left|\frac{a_1}{a_0z}\right|+\left|\frac{a_2}{a_0z^2}\right|+\cdots+\left|\frac{a_n}{a_0z^n}\right|$, 并且当 z 的模增大时, 这个和的每一项就减小, 因此这个和也就减小. 所以对于模大于某一数 R' 的一切 z 来说, 这个括号的模就小于 $\frac{1}{2}$.

但是对于一切使得括号 $\Omega=\left[1+\left(\frac{a_1}{a_0z}+\frac{a_2}{a_0z^2}+\cdots+\frac{a^n}{a_0z^n}\right)\right]$的模大于1/2 的 z 来说, 第一个因子 a_0z^n 的模$|a_0|\cdot|z|^n$随着 z 的模的增大而可以任意增大, 所以无论给出一个多么大的数 G, 总存在一个正数 R 使得对于模大于 R 的一切 z 来说, $|f(z)|=|a_0|\cdot|z|^n\cdot|\Omega|$大于 G.

曲面 M 的极小值的存在 如果曲面 M 上与复平面上的点 α 对应的点的立坐标的值小于或等于点 α 的某个邻域内一切点(即以 α 为心的很小的圆内的所有点)所对应的值, 我们就说曲面 M 在复平面的点 α 处有极小值.

设与复平面的原点(即 $z=0$ 这一点)对应的曲面 M 上的点的立坐标是 g, 即$|f(0)|=g$. 我们取 $G>g$. 曲面 M 的一切点的立坐标 t 是非负的, 且当复平面的 z 连续移动时 t 也连续地变动. 又有一 R 使曲面 M 在以 R 为半径以原点为心的圆外有立坐标 $t>G$, 而在圆心处有立坐标 $t=g<G$. 达朗贝尔认为这一事实包含有下列的显然推论: 即在这个圆内总有一点使得在它那里的立坐标最小, 确切地说, 在它那里的曲面 M 上的点的立坐标 t 小于或等于圆内其他一切点的立坐标, 即曲面 M 至少有一极小值.

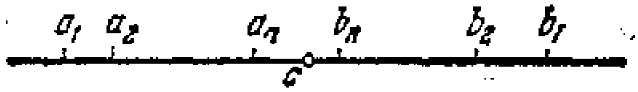


图 4

这个极小值的存在的严格证明基于下列的关于实数集的连续性公理.

如果给定两个实数串: $a_1\leqslant a_2\leqslant\cdots\leqslant a_n\leqslant\cdots$ 及 $b_1\geqslant b_2\geqslant\cdots\geqslant b_n\geqslant\cdots$, 使得对一切 n 都有 $b_n>a_n$, 且当 $n\to\infty$ 时 $b_n-a_n\to 0$, 那末有且仅有一个实数 c 存在使得对一切 n 都有 $a_n\leqslant c\leqslant b_n$.

连续性所揭示的几何性质是: 如果在直线上给定一个线段串$[a_n, b_n]$(图4)使得每一后面的线段包含在前面的里面, 并且线段的长可变到任意小, 那

末有一点 c 存在使得它属于这个线段串的一切线段内，换句话说，线段"收缩"到某一点，而不是"收缩"到"空地"。

由于线段$[a_n, b_n]$之长当 n 增加时趋近于零，这样的点只有一个。从这个关于实数集(即数轴上一切点的集)的连续性质出发立刻得到关于复数(即平面上的点)的连续性质。我们给出这个性质的几何说法。

如果在平面给一串边与坐标轴平行的长方形串 $\Delta_1, \Delta_2, \cdots, \Delta_n, \cdots$，它们中间每一个后面的都包含在前面的里面，并且它们的对角线无限地减小，那末有且仅有一点是属于这串长方形中的每一个。这个平面的连续性质由关于直线的连续性质直接得出。为了证明它，只须将长方形投射到坐标轴上就够了。

现在容易建立所谓的波尔察诺-魏尔斯特拉斯定理。

如果在某一长方形里给出了无限的点串 $z_1, z_2, \cdots, z_n, \cdots$，那末在长方形的里面或边界上有点 z_0 存在使得在它的任何一个邻域(即以 z_0 为心的任意充分小的圆)里有点串 $z_1, z_2, \cdots, z_n, \cdots$ 中的无穷多个点。

为了证明这个定理，我们用 Δ_1 表示给定的长方形，用平行于坐标轴的直线将它分为四个相等的部分，在这四部分中至少有一部分包含给定的点串中的无穷多个点，用 Δ_2 记这一部分。将 Δ_2 再分为四个相等的部分，并且取其中含有给定点串中无穷多个点的那一部分为 Δ_3，等等。

我们得到了对角线无限地减小的一个长方形串，并且它们是一个套一个的。按照连续性质有一点 z_0 属于一切作出的长方形。它就是所要求的。事实上，无论取 z_0 的怎样小的邻域，长方形串 $\Delta_1, \Delta_2, \cdots, \Delta_n, \cdots$ 必从某一个开始整个地在这个邻域里面，这仅仅只要对角线小于邻域的半径就够了。而每一长方形都含有点串 $z_1, z_2, \cdots, z_n, \cdots$ 中的无穷多个点。这样波尔察诺-魏尔斯特拉斯定理就被证明了。

现在容易证明关于多项式的模 $|f(z)|$ 的极小值的定理。同前面一样，设 $|f(0)|=g$; G 是某一大于 g 的数; R 是使得当 $|z|>R$ 时，$|f(z)|>G$ 的数。

若 $g=0$，即 $f(0)=0$，那末多项式的模 $|f(z)|$ 在点 0 处有极小值，这是因为在一切点上都 $\geqslant 0$。

若 $g>0$，且对一切点 z 都有 $|f(z)|\geqslant g$，那末 $|f(z)|$ 也在点 0 有极小值，设 $g>0$ 且存在点 z 使 $|f(z)|<g$; 此时在数列

$$0, \frac{g}{n}, \frac{2g}{n}, \cdots, \frac{ng}{n}=g \qquad (^*)$$

中，有一最大的 $c_n=\frac{ig}{n}$ 使一切的 $|f(z)|\geqslant c_n$，而对于列 $(^*)$ 中的后一数 $c_n'=\frac{i+1}{n}g$，就至少有一点 z_n 使 $|f(z_n)|<c_n'$。

我们使 n 无限地增大．对于一切 n 都有 $|z_n|\leq R$，因为如果 $|z_n|>R$，那末 $|f(z)|$ 就大于 G，因而大于 g．

这样一来，一切点 z_n 都在以坐标原点为心而边长为 $2R$ 的长方形内．很可能这些点的某些个有相互重合的现象．

按照波尔察诺-魏尔斯特拉斯定理有点 z_0 存在，使得在它的任何一个邻域里有点邑 $z_1, z_2, \cdots, z_n, \cdots$ 中的无穷多个点．

我们来确定恰恰在点 z_0 处 $|f(z)|$ 有极小值．

事实上，设 z 是任一点．此时

$$|f(z)|\geq c_n=c_n'-\frac{g}{n}>|f(z_n)|-\frac{g}{n}$$

$$=|f(z_0)|+(|f(z_n)|-|f(z_0)|)-\frac{g}{n}.$$

这个不等式对任何 n 都是成立的，如果取一 n 使得 z_n 充分地接近于 z_0，那末由 $|f(z)|$ 的连续性得出 $|f(z_n)|-|f(z_0)|$ 的绝对值与 $\frac{g}{n}$ 一样可以任意地小．

因此 $|f(z)|\geq|f(z_0)|$，即 $|f(z)|$ 在点 z_0 处达到了极小值．

达朗贝尔引理　由于曲面 M 的一切点的立坐标 t 是非负的，那末多项式 $f(z)$ 的根（即使得 $f(z)$，因而也使它的模 $|f(z)|$ 等于零的复平面的点 z）显然对应于模曲面 M 的一个极小值．然而如达朗贝尔所证明的，反过来也是正确的，即曲面 M 的每一极小值都接触到复平面，即在极小值处有多项式 $f(z)$ 的根．换句话说，曲面 M 没有一个极小值是正的而不是零．这由所谓达朗贝尔引理推出：

若 α 是任一给定的使得 $f(\alpha)\neq 0$ 的复数，则总可以找到一个绝对值任意小的复数 h 使 $|f(\alpha+h)|<|f(\alpha)|$．

证　我们考虑多项式

$$f(\alpha+h)=a_0(\alpha+h)^n+a_1(\alpha+h)^{n-1}+\cdots+a_{n-1}(\alpha+h)+a_n,$$

并且将这一具有两个文字 α 及 h 的多项式按 h 的升幂排列．在这个多项式中不含 h 的项

$$a_0\alpha^n+a_1\alpha^{n-1}+\cdots+a_{n-1}\alpha+a_n=f(\alpha)\neq 0,$$

这是因为我们假设 $f(\alpha)\neq 0$．又由于我们假设 $a_0\neq 0$，因此有含 h^n 的项 a_0h^n，至于具有 h 的中间方幂的那些项可能一部分或全部不

出现．设在这种写法中所出现的 h 的最低次幂是 m 次的，此处 $1\leqslant m\leqslant n$，于是这个写法具有形式

$$f(\alpha+h)=f(\alpha)+Ah^m+Bh^{m+1}+Ch^{m+2}+\cdots+a_0h^n.$$

将它写成

$$f(\alpha+h)=f(\alpha)+Ah^m+Ah^m\left(\frac{B}{A}h+\frac{C}{A}h^2+\cdots+\frac{a_0}{A}h^{n-m}\right),$$

式中 $A\neq0$，而 B，C 等数可能等于零也可能不等于零．

经过这些准备以后，达朗贝尔引理的证明粗略地说来是这样进行的．取 h 是一个使向量 Ah^m 的长小于向量 $f(\alpha)$ 的长而模充分小的复数；并且它具有使向量 Ah^m 的方向与向量 $f(\alpha)$ 的方向相反的辐角，此时向量 $f(\alpha)+Ah^m$ 就比向量 $f(\alpha)$ 短．但是对模充分小的一切 h 来说，括号 $\left(\frac{B}{A}h+\frac{C}{A}h^2+\cdots+\frac{a_0}{A}h^{n-m}\right)$ 的模可以任意小，比方说小于 1，于是向量 $\Delta=Ah^m\left(\frac{B}{A}h+\frac{C}{A}h^2+\cdots+\frac{a_0}{A}h^{n-m}\right)$ 的长就比 Ah^m 的长要短．所以象在图 5 中所看见的那样，哪怕是向量 Δ 恰恰与向量 Ah^m 有相反的方向，向量 $f(\alpha+h)=f(\alpha)+Ah^m+\Delta$ 同样是比向量 $f(\alpha)$ 短．

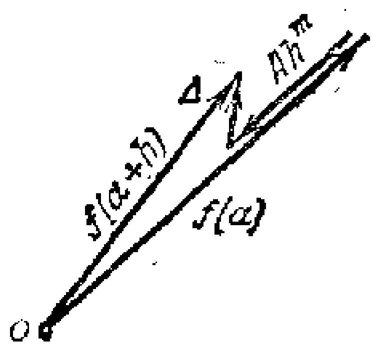


图 5

这个证明的详细情形如下所述：

1．因为对应于复数的乘法是因子的辐角相加，故 h 的辐角必须满足等式：

$$\arg A+m\cdot\arg h=\arg f(\alpha)+180^\circ,$$

即

$$\arg h=\frac{\arg f(\alpha)-\arg A+180^\circ}{m}.$$

2．括号 $\left(\frac{B}{A}h+\frac{C}{A}h^2+\cdots+\frac{a_0}{A}h^{n-m}\right)$ 的模不大于它的各项的模的和 $T=\left|\frac{B}{A}h\right|+\left|\frac{C}{A}h^2\right|+\cdots+\left|\frac{a_0}{A}h^{n-m}\right|$，并且当 h 的模任意

减小时，这个和的每一项可以任意减少，因而这个和也可以任意减少. 所以如果 h 是有上述辐角的复数，而 h_0 是使得模小于 h_0 的一切 h 都满足条件 $|Ah^m|<|f(\alpha)|$ 及 $T<1$ 的数，那末对于一切具有这个辐角而模小于 h_0 的 h 都有 $|f(\alpha+h)|<|f(\alpha)|$，这就证明了达朗贝尔引理.

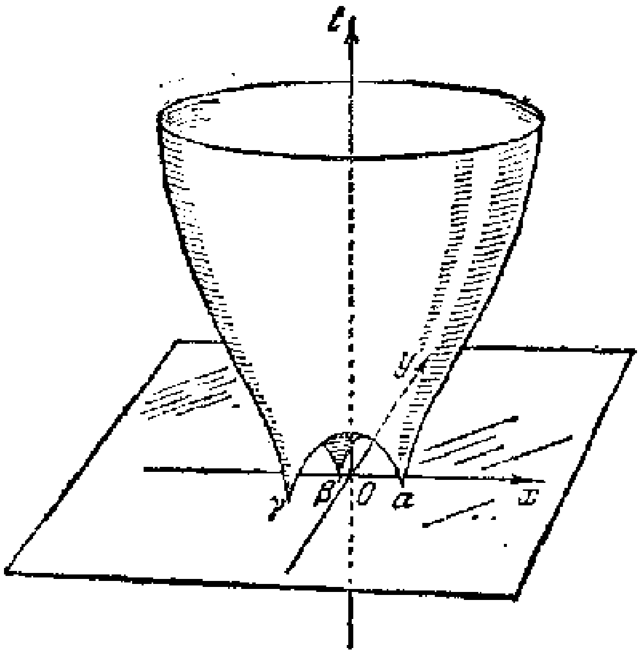


图 6

由达朗贝尔引理直接得出多项式 $f(z)$ 的模的曲面 M 的每一极小值给出它的多项式的根. 事实上，如果在点 α 处有 $f(\alpha)\neq 0$，那末由达朗贝尔引理在与它任意接近之处有点 $\alpha+h$ 使 $|f(\alpha+h)|<|f(\alpha)|$，即不存在以点 α 为心的圆使在它内部的一切点上，$f(z)$ 的模都不小于 $f(\alpha)$ 的模，所以 $f(z)$ 的模不可能在点 α 处取极小值，这样高等代数的基本定理就被证明了.

模曲面 M 的一般形式 多项式 $f(z)$ 的模曲面 M 位于复平面 z 之上，它具有图 6 所指出的形式，能够证明曲面在较高的 t 处与由 n 阶抛物线 $t=|a_0|x^n$ 绕 Ot 轴所得的曲面很少不同. 但对于小的 t 来说，曲面 M 有极小值，极小值的个数等于方程 $f(z)=0$ 的不同的根数. 曲面 M 由所有这些极小值在复平面 z 上支起来.

§4. 多项式的根在复平面上的分布的研究

对于实用很重要的一系列问题是与这样的问题有关联的，即不解出方程而得到关于它的根在复平面上的分布的这些或那些知识，第一个这样的问题是关于确定方程的实数根的个数的问题，就是说，如果给了一个实系数方程，那末不解出它而按照与它的系数

有关的某种判别条件来指出它是否有实数根，如果有，有多少；或者有多少正根和有多少负根；或者它在给定的界 a 及 b 之间有多少个实数根.

多项式的导数 多项式的导数在本段中担负着主要的作用. 什么是给定的函数的导数已在第二章中阐明.

如大家所知，多项式 $a_0x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n$ 的导数等于多项式 $na_0x^{n-1}+(n-1)a_1x^{n-2}+\cdots+a_{n-1}$.

在第二章中仅仅考虑过实变数函数的导数的概念，在代数里必须认为变数取任意复数值并且引入复系数多项式加以考虑.

然而导数的定义仍然可以保留以前的定义而认为是函数的增量与自变数增量之比的极限，复系数多项式导数计算的公式及基本的微分法则（和、乘积与方幂的微商）仍然如前[1].

多项式的单根和重根 在这一章的 §2 里曾经断定如果数 a 是多项式 $f(x)$ 的根，那末 $f(x)$ 能被 $x-a$ 整除，如果 $f(x)$ 不能被 $(x-a)^2$ 整除，那末数 a 就叫做多项式 $f(x)$ 的单根，一般地，如果多项式 $f(x)$ 能被 $(x-a)^k$ 整除而不能被 $(x-a)^{k+1}$ 整除，那末数 a 叫做 k 重根.

k 重根 a 常常看作是 k 个相等的根. 这种看法的理由是在 $f(x)$ 分解为一次因子的式子中，因子 $(x-a)^k$ 是 k 个相同的因子 $(x-a)$ 的乘积.

如果约定每一个根所算的次数就是它的重数，那末由于每一个 n 次多项式分解为 n 个一次因子的乘积，多项式的根数就等于次数.

下列的定理是正确的：

1. 多项式的单根不是它的导数的根.

2. 多项式的重根是它的导数的根而重数少一.

事实上，设 $f(x)=(x-a)^kf_1(x)$. 并且 $f_1(x)$ 不被 $x-a$ 整除，即 $f_1(a)\neq0$, 则

1) 参看第二卷第九章.

$$\begin{aligned}f'(x) &= k(x-a)^{k-1}f_1(x)+(x-a)^k f_1'(x)\\ &= (x-a)^{k-1}[kf_1(x)+(x-a)f_1'(x)]\\ &= (x-a)^{k-1}F(x).\end{aligned}$$

多项式 $F(x)=kf_1(x)+(x-a)f_1'(x)$ 不能被 $x-a$ 整除，这是因为 $F(a)=kf_1(a)\neq 0$。

因此当 $k=1$ 时 $f'(x)$ 不能被 $x-a$ 整除，而当 $k>1$ 时，$f'(x)$ 能被 $(x-a)^{k-1}$ 整除，但不能被 $(x-a)^k$ 整除，于是这两个定理获证。

罗尔定理及它的某些推论 由习知的罗尔定理[1]知如果实数 a 与 b 是实系数多项式的根，那末在 a 与 b 之间有一实数 c 存在，而它是导数的根。

由罗尔定理产生下列有趣的推论。

1．如果多项式 $f(x)=a_0x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n$ 的一切根都是实的，那末它的导数的一切根也是实的，在 $f(x)$ 的相邻两根之间有 $f'(x)$ 的一个根并且是一个单根。事实上，设 $x_1<x_2<\cdots<x_k$ 是 $f(x)$ 的根，它们分别具有重数 $m_1, m_2, \cdots, m_k$。显然 $m_1+m_2+\cdots+m_k=n$。

此时按照关于重根的定理，导数 $f'(x)$ 有根 $x_1, x_2, \cdots, x_k$，其重数分别是 $m_1-1, m_2-1, \cdots, m_k-1$；又由罗尔定理还至少有根 $y_1, y_2, \cdots, y_{k-1}$ 分别在 $f(x)$ 的相邻两根的区间 $(x_1, x_2), (x_2, x_3), \cdots, (x_{k-1}, x_k)$ 里。这样 $f'(x)$ 的实根的个数（考虑到重数）至少等于 $(m_1-1)+(m_2-1)+\cdots+(m_k-1)+k-1=n-1$。但 $f'(x)$ 是 $n-1$ 次多项式，所以有 $n-1$ 个根（考虑到重数），因此 $f'(x)$ 的一切根都是实的，$y_1, y_2, \cdots, y_{k-1}$ 是单根，并且除了 $x_1, x_2, \cdots, x_k$ 及 $y_1, y_2, \cdots, y_{k-1}$ 以外，$f'(x)$ 没有其他的根。

2．如果多项式 $f(x)$ 的一切根都是实的，并且其中有 p 个（考虑到重数）是正的，那末 $f'(x)$ 有 p 个或 $p-1$ 个正根。

事实上，设 $x_1<x_2<\cdots<x_k$ 是多项式 $f(x)$ 的一切正根，他们分别具有重数 $m_1, m_2, \cdots, m_k$，则 $m_1+m_2+\cdots+m_k=p$。导数

1）这个定理是中值定理的最简单的情形，已经在第二章第 138 页谈到它。

$f'(x)$ 将有下列正根：$x_1, x_2, \cdots, x_k$ 分别具有重数 $m_1-1, m_2-1, \cdots, m_k-1$；$y_1, y_2, \cdots, y_{k-1}$ 分别是位于区间 $(x_1, x_2), \cdots, (x_{k-1}, x_k)$ 内的单根，很可能还有一个单根 y_0 位于区间 (x_0, x_1) 之内，此处 x_0 是 $f(x)$ 的最大的非正根，因此 $f'(x)$ 的正根数等于 $(m_1-1)+\cdots+(m_k-1)+k-1=p-1$ 或 $(m_1-1)+\cdots+(m_k-1)+(k-1)+1=p$，这就是所要证明的。

笛卡儿符号定则 笛卡儿在 1637 年发表的著名的《几何学》一书中第一次阐述了解析几何，同时顺便给出了第一个著名的代数定理，这就是关于多项式的根在复平面上的分布理论的"笛卡儿符号定则"，它可以表述如下：

如果方程的系数是实数并且知道它的一切根都是实的，那末它的正根的个数在将重数计算在内时就等于它的系数序列的变号数，如果它有复根，那末正根数就等于这个变号数或比这个变号数小某一个偶数。

首先我们解释方程的系数序列的变号数，为了得到这个数，将方程的一切系数按未知数的降幂顺序写下，其中包括 x^n 的系数及常数项，但去掉一切等于零的系数。再考察所得到的序列的一切相邻的数的数对，如果在这样的数对中，两数的符号不同，那末就叫做一个变号，例如，如果给了方程

$$x^7+3x^5-5x^4-8x^3+7x+2=0,$$

那末它的系数的序列

$$1,\ 3,\ -5,\ -8,\ 7,\ 2$$

有两个变号。

现在我们来证明定理的第一部分[1]。

不失其一般性可以认为多项式 $f(x)=a_0x^n+\cdots+a_n$ 首系数 a_0 是正的。

我们首先断定：如果 $f(x)$ 仅有实根，并且有 p 个(计算重数)是正的，那末 $(-1)^p$ 是 $f(x)$ 的最后一个不等于零的系数的符号。

事实上，设

1) 可以给出不依赖于导数的另外的直接证明，但是证明稍长一点。

$$f(x)=a_0x^n+\cdots+a_kx^{n-k}$$
$$=a_0x^{n-k}(x-x_1)\cdots(x-x_p)(x-x_{p+1})\cdots(x-x_{n-k}),$$

式中 x_1, …, x_p 是 $f(x)$ 的正根, x_{p+1}, …, x_{n-k} 是 $f(x)$ 的负根; 而每一个根所算的次数就是它的重数, 此时 $a_k=a_0(-1)^px_1\cdots x_p(-x_{p+1})\cdots(-x_{n-k})$, 而由于 a_0, x_1, …, x_p, $-x_{p+1}$, …, $-x_{n-k}$ 是正数, a_k 的符号是 $(-1)^p$[1].

今后的证明是应用数学归纳法.

对于一次多项式定理是显然的, 事实上, 一次多项式 a_0x+a_1 有唯一的根 $-\frac{a_1}{a_0}$, 它仅在 a_0 与 a_1 有相反符号的情形才是正数.

现在假设定理对实系数的一切 $n-1$ 次多项式已经证明, 在这个假设下, 我们对 n 次多项式 $f(x)=a_0x^n+\cdots+a_{n-1}x+a_n$ 来证明它.

1. $a_n=0$, 考虑多项式 $f_1(x)=a_0x^{n-1}+\cdots+a_{n-1}$. 多项式 $f(x)$ 与 $f_1(x)$ 的正根是相同的; 而它们的系数序列的变号数也是相同的, 对于多项式 $f_1(x)$, 笛卡儿法则是正确的, 因此对于多项式 $f(x)$, 它也是正确的.

2. $a_n\neq 0$, 我们来考虑导数 $f'(x)=na_0x^{n-1}+(n-1)a_1x^{n-2}+\cdots+a_{n-1}$.

显然当 a_n 与导数的最后一个不等于零的系数的符号相同时, 多项式的系数序列的变号数与导数 $f'(x)$ 的相同; 符号相反时, 就多一个.

由在证明开始时所指出的事实知在第一种情形下 $f(x)$ 与 $f'(x)$ 的正根数有相同的奇偶性, 而在第二种情形下则相反. 但是象我们在罗尔定理中所见到的, 多项式(如果它的一切根都是实的)的正根数与它的导数的正根数相等或者多一个, 利用这一点, 我们断言在第一种情形下 $f(x)$ 的正根和 $f'(x)$ 的一样多, 而在第二种情形下就多一个. 由归纳假设知, 笛卡儿定则对于 $f'(x)$ 是正

---

1) 注意这个结论对 $f(x)$ 的根中有复数的情形仍然正确.

确的，即 $f'(x)$ 的正根数等于它的系数序列的变号数，因此在两种情形下 $f(x)$ 的正根数等于系数序列的变号数，这就是所要证明的.

笛卡儿定则的第二部分的建立并不复杂，但我们不再在此给出它的证明.

注 1 笛卡儿定理的第一个论断是特别重要的，因为在很多的实际问题中明明知道所得到的方程的一切根都是实的. 在这些情况下能够很快地知道多少个是正根，多少个是负根，而方程有多少个根是零则可立刻知道.

注 2 如果在所考虑的多项式中令 $x=y+a$，式中 a 是一个任意给定的数，即写出多项式 $f(y+a)$，那末这个多项式的正根 y 是且仅仅是由给定的多项式 $f(x)$ 的大于 a 的那些根 x 所得到的，所以一切根都是实数的多项式 $f(x)$ 的在 a 与 $b(b>a)$ 间的根的个数等于多项式 $f(x+a)$ 的变号数减去多项式 $f(x+b)$ 的变号数；如果 $f(x)$ 的根不全是实数，那末可以证明它等于这个差或者比这个差小一个偶数. 这就是所谓的比由当定理.

斯图谟定理 然而笛卡儿符号定则以及比由当定理并没有给出下述问题的答案，这些问题是：一个给定的实系数方程是不是有实根. 它总共有多少个实根，并且在给定的限 a 与 b 之间它有多少个实根，在两个多世纪的时间内，数学家们尝试着解决这些问题，然而毫无结果，在这个方向下有一大串结果，它们是属于笛卡儿、牛顿、比由当、西尔维斯特尔、富里哀及其他数学家的，但是直到 1835 年法兰西数学家斯图谟提出解决这三个问题的方法以前，他们甚至连第一个问题也解决不了.

斯图谟导出结论的方法其实并不复杂，但它毕竟是长久以来所寻求的然而却没找到的方法，斯图谟本人很幸运地解决了这个对实用来说是著名的并且非常重要的代数问题. 当他在讲演中阐述到自己的这些结果时，他常说："这就是我负担了它们的名字的那些定理." 应该指出斯图谟解决这些问题并非偶然；他很多年都在思考与它们相邻近的问题.

设 $f(z)$ 是一个具有实系数且没有重根的多项式[1)]，$f_1(z)$ 是它的导数 $f'(z)$. 用 $f_1(z)$ 除多项式 $f(z)$，并用 $f_2(z)$ 表示出这个除法所得到的余式反号后的多项式. 然后，用 $f_2(z)$ 除 $f_1(z)$，并用 $f_3(z)$表余式反号后的多项式，等等.

能够证明在所作的序列中，最后一个不等于零的多项式$f_s(z)$是一个常数 c.

斯图谟定理是这样的：如果 $a<b$ 是两个实数，并且不是 $f(z)$ 的根，那末将 $z=a$ 及 $z=b$ 代入多项式

$$f(z),\ f_1(z),\ \cdots,\ f_{s-1}(z),\ c$$

中，我们得到两串实数

$$f(a),\ f_1(a),\ \cdots,\ f_{s-1}(a),\ c, \tag{I}$$

$$f(b),\ f_1(b),\ \cdots,\ f_{s-1}(b),\ c, \tag{II}$$

于是序列(I)的变号数大于或等于序列(II)的变号数，而这两个变号数之差恰恰等于$f(z)$在 a 与 b 间的实根数，换句话说，a 与 b 间的实根数等于序列(I)由 a 变到 b 时所丢掉的变号数.

斯图谟定理的证明并不比笛卡儿定理的证明要难，然而我们不在此处给出.

斯图谟定理给出了计算实系数多项式在实轴上的任何区间内的根数，所以对于任一给定的多项式，应用斯图谟定理足可以查清多项式在实轴上的根的分布的详细情形，特别是根的分离，即求出一组区间，使得每一区间中只含有多项式的一个根.

在很多应用中，对于多项式的复根解决类似的问题有不小的意义，由于复数不能用直线上的点来表示，而能用平面上的点来表示，因此不能说包含在“区间”里的复根；而用区域代替区间来考虑，即考虑用这种或那种所分出的平面的一部分.

因此在应用中提出下列关于复根的问题：

---

1) 原文中没有$f(z)$没有重根这一假定，但是如果没有这一假定，那末多项式串 $f(z)$, $f_1(z)$, $f_2(z)$, $\cdots$ 中最后一个不等于零的多项式 $f_s(z)$便不是常数，因此译者在此添上这一假定，至于在这一假定去掉的情况下，斯图谟定理仍然正确，只是将 c 换成 $f_s(z)$罢了. ——译者注

给了多项式 $f(z)$ 及复平面上的一个区域，需要知道多项式在这个区域里的根的个数.

我们将假定区域是由一封闭的围道包住的(图 7)，并且在区域的围道上多项式 $f(z)$ 没有根.

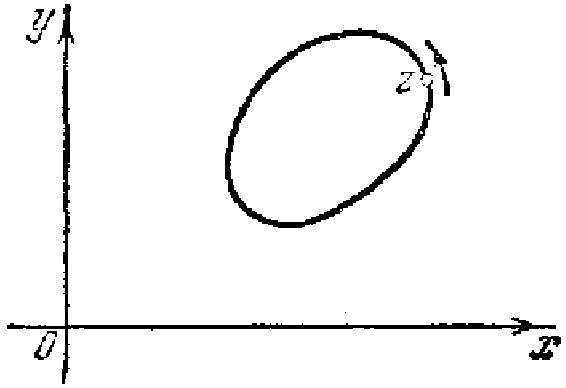


图 7

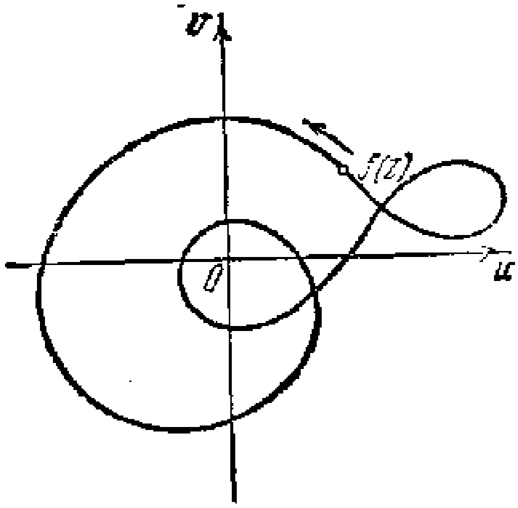


图 8

设想点 z 沿区域围道的正方向通过一次，而多项式的每一个值也能用平面的点来表示，当 z 连续地变动时多项式 $f(z)$ 也连续地变动，因此在 z 绕区域围道一圈时，$f(z)$ 就描出一条封闭的曲线，根据假设知 $f(z)$ 在围道上的任何一点都不是零，所以这条曲线不通过坐标原点(图 8).

下面的定理便是上面所提出的问题的答案：

辐角原理 多项式 $f(z)$ 在一由封闭的围道 C 包住的区域内的根数等于当 z 沿 C 的正方向通过一次时 $f(z)$ 绕坐标原点的圈数.

为了证明这个定理，我们将 $f(z)$ 分成一次因子的乘积

$$f(z)=a_0z^n+a_1z^{n-1}+\cdots+a_n=a_0(z-z_1)(z-z_2)\cdots(z-z_n).$$

我们知道若干个复数的乘积的辐角等于因子的辐角的和，因此

$$\arg f(z)=\arg a_0+\arg(z-z_1)+\arg(z-z_2)+\cdots+\arg(z-z_n).$$

我们用 $\Delta\arg f(z)$ 表示 z 通过围道 C 时 $f(z)$ 的辐角的增量. 显然 $\Delta\arg f(z)$ 是 2π 的一个倍数，这个倍数就是点 $f(z)$ 绕原点的

次数.

显然

$$\Delta\arg f(z)=\Delta\arg a_0+\Delta\arg(z-z_1)+\Delta\arg(z-z_2)$$
$$+\cdots+\Delta\arg(z-z_n).$$

因为 a_0 是一个常数，$\Delta\arg a_0=0$. 又 $z-z_1$ 可以用点 z_1 到点 z 的向量来表示. 假设 z_1 位于区域的内部，从几何来看，当点 z 通过 C 时，向量 $z-z_1$ 以 z_1 为原点绕一整周(图 9)，因此 $\Delta\arg(z-z_1)=2\pi$. 现在假设点 z_2 位于区域之外，在这种情形下，向量在移动后回到原来的位置，但并没有绕原点一周，因此 $\Delta\arg(z-z_2)=0$. 用这样的方法可以考察一切根. 因此 $\Delta\arg f(z)$ 等于 2π 乘以 $f(z)$ 的位于区域内部的根数. 故 $f(z)$ 的位于区域内部的根数等于点 $f(z)$ 绕原点的次数，这就是所要证明的.

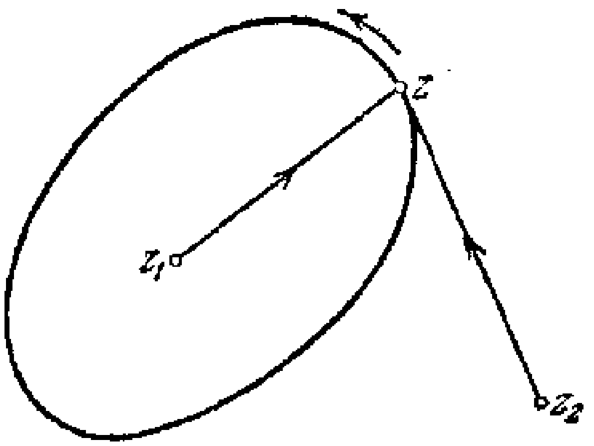


图 9

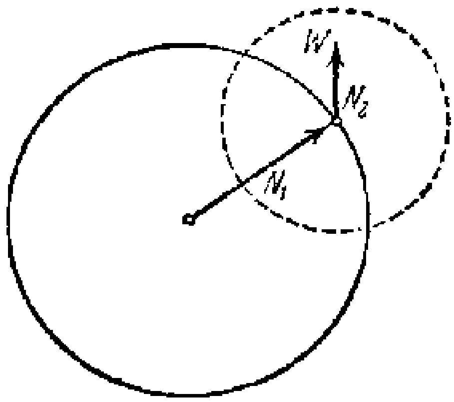


图 10

以上所证明的定理给出了解决所提出的问题的每一特殊情形的方法，就是说要精确地画出点 $f(z)$ 所描出的曲线，为了这一点，必须在围道 C 上取一组相当稠密的点 z，计算出与它们相应的 $f(z)$ 的值，并且用连续的曲线把它们联结起来，然而在某些情况下，可能不必进行这些令人厌倦的计算，我们举出一个数字作为例子来说明这一点.

例 求出多项式 $f(z)=z^{11}+5z^3-2$ 在以一为半径以原点为心的圆内的根数.

在所给出的圆周 $|z|=1$ 上，构成多项式 $f(z)$ 的三项之中有一

项(即 $5z^2$)的绝对值大于其余各项的绝对值. 事实上 $|5z^2|=5$, 而 $|z^{11}-2|\leqslant|z|^{11}+2=3$, 这种情况使我们能象下面那样来讨论, 我们以 ω 表 $z^{11}+5z^2-2$, 以 N_1 表 $5z^2$, 以 N_2 表 $z^{11}-2$. 当 z 绕单位圆一周时, $N_1=5z^2$ 就绕行半径是 5 的圆周两次, 这是因 $|N_1|=5$, $\arg N_1=2\arg z$. 而点 w 与 N_1 是由一个长度不超过 $|N_2|\leqslant 3$ 的向量"联系"着, 即点 w 到点 N_1 的距离总是小于点 N_1 到坐标原点的距离.

因此, 不论点 w 怎样绕着 N_1 "转"(图 10), 它都不能"独立"地绕原点转, 于是它绕原点的次数与点 N_1 绕原点的次数一样, 即绕原点两次. 因此 $f(z)$ 在所考虑的区域内的根数是 2.

古尔维茨问题 在力学中, 即在振动理论及调节理论中, 鉴别给定的多项式 $f(z)=a_0z^n+a_1z^{n-1}+\cdots+a_n$ (实系数)是否一切根都具有负的实数部分是很重要的, 即是否一切根在虚轴的左半平面上.

解决这个问题的一个判别条件很容易由辐角原理导出的一个想法得到.

我们设 $a_0>0$.

设点 z 沿虚轴从上到下地变动, 即设 $z=iy$, 并且 y 由 $+\infty$ 变到 $-\infty$(图 11), 则 $f(z)$ 描绘出某一有无穷远端点的曲线. 为了研究时技巧上更方便些, 我们将这一条曲线与函数

$$\begin{aligned}f_1(z)&=(i)^{-n}f(z)\\&=a_0y^n-a_2y^{n-2}+a_4y^{n-4}+\cdots-i(a_1y^{n-1}-a_3y^{n-3}+\cdots)\\&=\varphi(y)-i\psi(y)\end{aligned}$$

所描绘的曲线紧密地联系起来, 在上式中

$$\varphi(y)=a_0y^n-a_2y^{n-2}+\cdots,$$
$$\psi(y)=a_1y^{n-1}-a_3y^{n-3}+\cdots.$$

因为 $\arg i=\dfrac{\pi}{2}$, 于是 $\arg f_1(z)=-\dfrac{n\pi}{2}+\arg f(z)$, 因而 $f(z)$ 与 $f_1(z)$ 的辐角的增量相同.

我们来计算当 z 从上到下通过虚轴时点 $f_1(z)$ 的辐角的增量.

设 $$f(z)=a_0(z-z_1)(z-z_2)\cdots(z-z_n),$$
则
$$\arg f_1(z)=\arg(a_0^{1-n})+\arg(z-z_1)+\arg(z-z_2)$$
$$+\cdots+\arg(z-z_n).$$
从几何来看，显然当 z_k 在右半平面时 $\arg(z-z_k)$ 的增量等于 π，当 z_k 在左半平面时 $\arg(z-z_k)$ 的增量等于 $-\pi$(图 11).

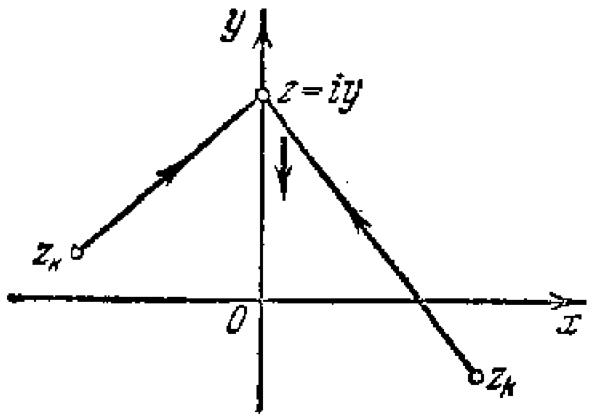


图 11

所以 $f_1(z)$ 的辐角的增量等于 $\pi(N_1-N_2)$，式中 N_1 是 $f(z)$ 在右半平面上的根数，N_2 是在左半平面上的根数. 因此使得一切在左半平面上的必要与充分条件是点 $f_1(z)$ 的辐角的增量等于 $-\pi n$，即点 $f_1(z)$ 沿着时针方向绕坐标原点转 n 个半圈(图 12).

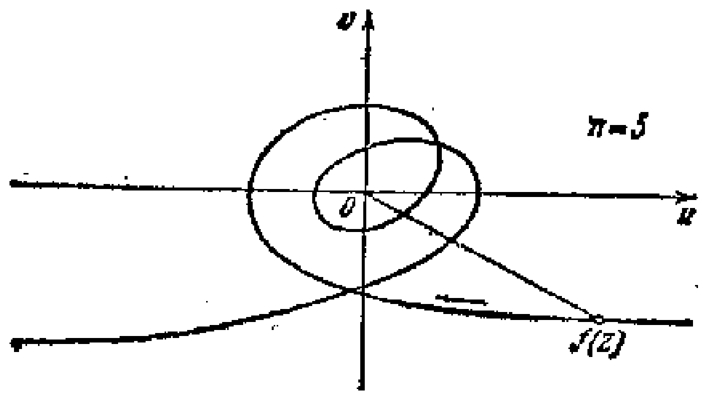


图 12

我们注意当 y 是 $\varphi(y)$ 的根时，点 $f_1(z)=\varphi(y)-i\psi(y)$ 与虚轴相交，当 y 是 $\psi(y)$ 的根时，与实轴相交. 由于 $\varphi(y)$ 的实根数不多于 n，$\psi(y)$ 的实根数不多于 $n-1$，从几何上来看，很容易确信 $f_1(z)$ 能够沿着时针方向转 n 个半圈当且仅当曲线从第四象限开始然后依次与虚轴的负的部分，实轴的负的部分，虚轴的正的部分，实轴的正的部分等等相交，并且与虚轴的交点(每一个半圈上恰有一点)总数等于 n，而与实轴的交点数等于 $n-1$(比半圈的个数少 1). 所以系数 a_1 应该是正的，而多项式 $\varphi(y)$ 及 $\psi(y)$ 的根应该都是实的并且互相间隔. 所谓互相间隔的意思是如果 $y_1>y_2>\cdots>y_n$ 是 $\varphi(y)$ 的根，并且按递降的顺序排

列，而 $\eta_1>\eta_2>\cdots>\eta_{n-1}$ 是 $\psi(y)$ 的根，那末 $y_1>\eta_1>y_2>\eta_2>\cdots>y_{n-1}>\eta_{n-1}>y_n$.

总之实系数多项式 $f(z)=a_0z^n+a_1z^{n-1}+\cdots+a_n$, $a_0>0$ 的一切根位于左半平面的充要条件是系数 a_1 是正的，并且多项式 $\varphi(y)=a_0y^n-a_2y^{n-2}+a_4y^{n-4}-\cdots$ 及 $\psi(y)=a_1y^{n-1}-a_3y^{n-3}+\cdots$ 的根都是实的且互相间隔.

这个条件与著名的古尔维茨条件等价，就是说行列式

$$a_1,\quad \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix},\quad \begin{vmatrix} a_1 & a_0 & a_{-1} \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{vmatrix},\quad \cdots,\quad \begin{vmatrix} a_1 & a_0 & a_{-1} & \cdots & a_{2-n} \\ a_3 & a_2 & a_1 & \cdots & a_{4-n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{2n-1} & a_{2n-2} & a_{2n-3} & \cdots & a_n \end{vmatrix}$$

都是正的，其中足码小于零或大于 n 的 a_i 都是零(关于此类行列式可参看第三卷第十六章 § 3).

§ 5. 根的近似计算法

斯图谟方法结合着界限的应用使我们可以进行实系数多项式的实根的“分离”，即对于每一这样的根，能够指出界限 a 及 b 来，而在它们之间仅有一个实根. 剩下的事就是要指出适当的方法能在线段 $a<b$ 上找出数 $\alpha_1<\alpha_2<\alpha_3<\cdots$ 及 $\beta_1>\beta_2>\beta_3>\cdots$，而他们是很快地愈来愈与所求的根接近，并且前者是它的不足近似值，而后者是过剩近似值. 两个近似值 α_k 及 β_k 中的每一个与所求根 x 之差显然小于它们的差 $\beta_k-\alpha_k$，这是因为所求的根是在它们之间. 于是如果讨论这些近似值，显然误差是小于它的.

多项式的图象 假设给定实系数 n 次多项式是

$$f(x)=a_0x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n.$$

我们来讨论在直角坐标中表示方程 $y=f(x)$ 的曲线，即这个多项式的图象. 这条曲线有时叫做 n 阶抛物线. 首先，对于任一实的 x，显然有一个且仅有一个确定的实数 $y=f(x)$，所以 f 的图象向左右延伸至任意远的地方，此外 $f(x)$ 及 $f'(x)$ 随 x 的连续变化而

连续变化，即没有跳跃，所以 f 的图象是光滑的曲线．对于绝对值充分大的 x 来说，第一项 a_0x^n 的绝对值大于其余一切项的和的绝对值，因为它们都具有较低的次数．由此推出：如果 n 是偶数而 $a_0>0$，那末 f 的图象就向左右方的高处无限伸展（如 $a_0<0$，就向低处）；如果 n 是奇数而 $a_0>0$，那末它就向右方高处无限伸展，而向左方低处伸展（如果 $a_0<0$，则反过来）．

f 的图象与 Ox 轴的交点，即使得 $y=f(x)=0$ 的点是对应于方程 $f(x)=0$ 的实根；它们的个数不大于 n．在图象 $y=f(x)$ 的极大值与极小值处，导数 $f'(x)=0$，因此极大值与极小值的总数不大于 $n-1$．如果在某一段里 $f''(x)>0$，那末在这里一阶导数是上升的，即图象本身是向上凹的；如果 $f''(x)<0$，那末它就向下凹了．如果已经知道 $f'(x)=0$ 的某一根是复的，f 的图象的极大值和极小值的个数就小于 $n-1$．

我们可以举多项式

$$f(x)=x^3-3x+1 \text{（图 13）},$$

$$f(x)=x^4-x^3-4x^2+4x+1 \text{（图 14）}$$

的图象作为例子．

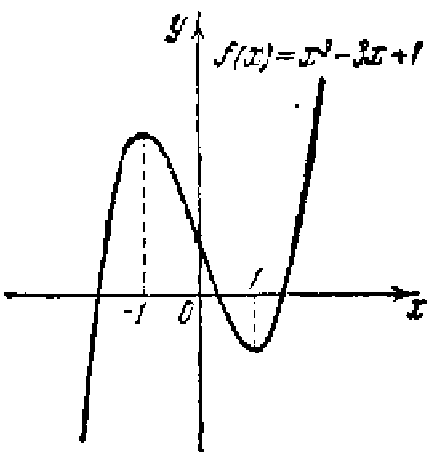


图 13

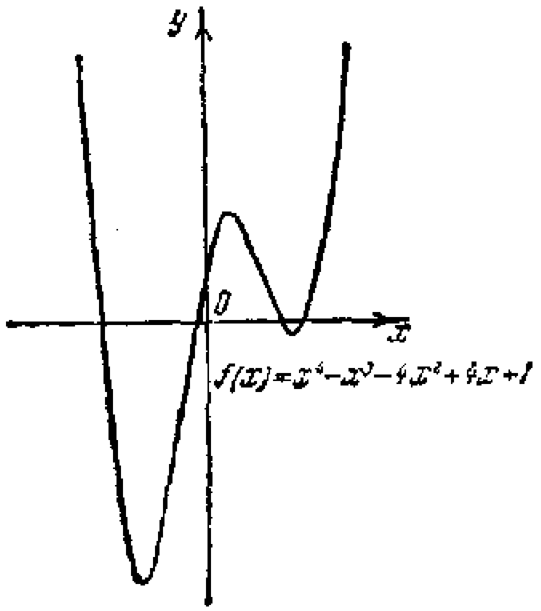


图 14

作出多项式的图象后，很容易找到它的根的近似值．因为根就是图象与 Ox 轴的交点的横坐标．

盈亏法 将任何一个有理整数(例如3)代入多项式 $f(x)$，然后代入4, 5, 6, ⋯. 如果4, 5, 6代入后所得结果与代入3的结果同号，而代入7后反号，那末显然在6与7之间至少有 $f(x)$ 的一个根. 现在我们代入6; 6.1; 6.2; ⋯，于是找到两个相邻的数，例如6.4及6.5，它们代入后给出了不同的符号. 我们就知道在它们之间至少有一根. 进一步以6.4; 6.41; 6.42; 6.43; ⋯代入并且仍然求出最靠近根的界，例如是6.42及6.43等等. 这个方法就叫做"盈亏法". 在对每一计算步骤加上适当的多项式变换后，这个方法还可以大大简化，使得除开第一步的每一步骤都不代入分数而代入整数，并且仅代入整数1, 2, 3, ⋯, 9. 但我们不停留在这些简化上面.

切线方法和弦方法 切线方法就是所谓的牛顿法，弦方法就是直线插入法，也叫做虚位法(*regula falsi*). 我们或者单独地使用，或者为了估计误差而合起来使用，假设 $a<b$ 并且在 a 与 b 之间仅有多项式 $f(x)$ 的一个实根 [$f(a)$ 与 $f(b)$ 反号]，此外二阶导数 $f''(x)$ 在 a 与 b 之间处处都具有相同的符号. 在这种情形下，f 的图象在 $x=a$ 及 $x=b$ 之间的一部分是四种形式之一(图15).

在情形 I 及情形 II 之下，图象上具有横坐标 a 的点的切线与 Ox 轴相交于具横坐标 α_1 的点，它位于所求的根与 a 之间. 求出横坐标 α_1 并且进一步考察图象上具有横坐标 α_1 的点的切线，我们可类似地找到点 α_2，它位于点 α_1 与根之间；更进一步可类似地求 α_3 等等，这样将得到很好的不足近似值. 由图上可以看出这些值很快地接近于所求的根.

在情形 III 及 IV 之下，应该反过来从横坐标 b 开始，则得到点 β_1, β_2, β_3, ⋯，即一些很好的过剩近似值. 至于四种情形中哪一种会发生，按照 $f(a)$, $f(b)$ 及 $f''(x)$, $a<x<b$ 的符号是很容易断定的.

因为曲线 $y=f(x)$ 上具有横坐标 a 的点的切线的方程是

$$y-f(a)=f'(a)(x-a),$$

于是它与 Ox 轴的交点的横坐标 α_1 由等式

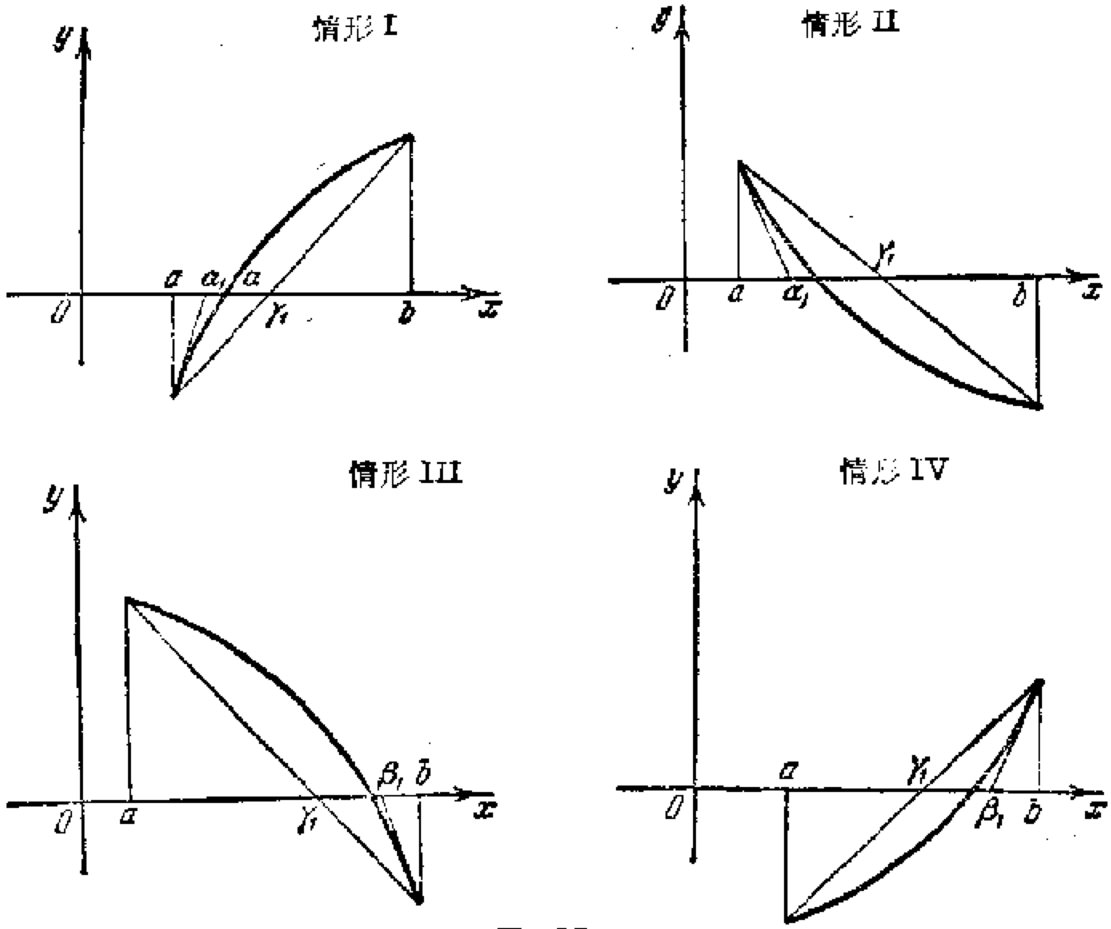


图 15

$$0-f(a)=f'(a)(\alpha_1-a)$$

得到，就是说
$$\alpha_1=a-\frac{f(a)}{f'(a)}.$$

进一步，$\alpha_2=\alpha_1-\dfrac{f(\alpha_1)}{f'(\alpha_1)}$，$\alpha_3=\alpha_2-\dfrac{f(\alpha_2)}{f'(\alpha_2)}$

等等.

类似地，

$$\beta_1=b-\frac{f(b)}{f'(b)},\quad \beta_2=\beta_1-\frac{f(\beta_1)}{f'(\beta_1)},\quad \beta_3=\beta_2-\frac{f(\beta_2)}{f'(\beta_2)}$$

等等.

这就是牛顿法[1].

1）从这些公式也可以给出上面由图形观察而得到的两个结论的严格证明，因为当 n 增大时，α_n（类似地对于 β_n）是单调地变化着，例如在情形 I 中，α_n 是递增并且有界的，由魏尔斯特拉斯引理知，α_n 趋近于极限 α，将这个公式中的 α_n 用 α 代入后就得到：$\alpha=\alpha-\dfrac{f(\alpha)}{f'(\alpha)}$，由此知 $f(\alpha)=0$，即 α 是多项式 f 的根。

直线插入法或者是虚位法有如下述：通过两个给定点的弦的方程和通过两点的直线的方程一样，有如下的形式：

$$\frac{x-a}{b-a}=\frac{y-f(a)}{f(b)-f(a)},$$

而它与 Ox 轴的交点的横坐标 γ_1 是由方程

$$\frac{x-a}{b-a}=\frac{0-f(a)}{f(b)-f(a)}$$

得到的，就是说

$$\gamma_1=-\frac{(b-a)f(a)}{f(b)-f(a)}+a=\frac{af(b)-bf(a)}{f(b)-f(a)}.$$

将它算作情形 I 及 II 的新的 b，而算作情形 III 及 IV 的新的 a 以后，我们对情形 I 及 II 就得到

$$\gamma_2=\frac{af(\gamma_1)-\gamma_1 f(a)}{f(\gamma_1)-f(a)},\quad \gamma_3=\frac{af(\gamma_2)-\gamma_2 f(a)}{f(\gamma_2)-f(a)}$$

等等.

在情形 III 及 IV 之下，我们将 γ_1 当作新的 a，就得到

$$\gamma_2=\frac{\gamma_1 f(b)-bf(\gamma_1)}{f(b)-f(\gamma_1)},\quad \gamma_3=\frac{\gamma_2 f(b)-bf(\gamma_2)}{f(b)-f(\gamma_2)}$$

等等.

这两个方法结合起来用是很重要的．因为(如图中所见)他们可以用来估计近似值的误差，显然，误差不会超过这两种近似值的差，这是由于所求的根在它们之间.

注意 很值得注意的是：$f(x)$ 是多项式而不是 x 的其他任何函数这一情况无论在牛顿法或直线插入法中都不起任何作用，即这两个方法以及它们的结合对于一切适合上述条件的超越方程都能采用.

罗巴切夫斯基方法 现今最通行的计算根(特别是复根)的方法之一是罗巴切夫斯基在他 1834 年发表的《代数》一书中提出的方法[1]，这个方法的基本想法伯努利还提出过.

首先我们指出：如果给定了一个多项式，它的根是 $\alpha_1, \alpha_2, \cdots,$

1) 这个方法是由唐代仑(1826)、罗巴切夫斯基(1834)及格雷菲(1837)独立地提出的.

x_n，那末容易写出一个以 $x_1^2, x_2^2, \cdots, x_n^2$ 为根的多项式，即写出以给定的多项式的根的平方为根的多项式，事实上，如果 $x_1, x_2, \cdots, x_n$ 是多项式

$$x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots a_n$$

的根，那末它等于

$$(x-x_1)(x-x_2)\cdots(x-x_n);$$

而多项式

$$x^n-a_1x^{n-1}+a_2x^{n-2}-\cdots\pm a_n$$

的根是给定的多项式的根的负数，它等于

$$(x+x_1)(x+x_2)\cdots(x+x_n).$$

因此这两个多项式的乘积是

$$(x^2-x_1^2)(x^2-x_2^2)\cdots(x^2-x_n^2),$$

因而只包含 x 的偶次幂项，如果在它里面令 $x^2=y$，那末就得到 y 的 n 次多项式

$$y^n+b_1y^{n-1}+b_2y^{n-2}+\cdots+b_n,$$

它等于

$$(y-x_1^2)(y-x_2^2)\cdots(y-x_n^2),$$

即它的根是 $x_1^2, x_2^2, \cdots, x_n^2$. 我们现在用列表方法将多项式

$$x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots+a_n$$

与多项式

$$x^n-a_1x^{n-1}+a_2x^{n-2}-\cdots\pm a_n$$

相乘而求出系数 b_k. 我们画一条线，在线上的一行顺次写下 1, $a_1, a_2, \cdots, a_n$；再在线下面对准每一系数 a_k，先写下它的平方 a_k^2，其次写出它的相邻两系数的乘积的负二倍

$$-2a_{k-1}a_{k+1},$$

然后写出对于 a_k 对称的两系数 a_{k-2}, a_{k+2} 的乘积的正二倍

$$+2a_{k-2}a_{k+2},$$

这样依次写出对于 a_k 对称的系数对的乘积的二倍，而符号是正负相间，直到有一边已没有系数为止，然后在 $a_1, a_2, a_3, \cdots, a_n$ 下面的各行的数前面依次加上符号 $-, +, -, \cdots\pm$，于是系数 b_k 就是写在线下的相应一行的数的和.

在得到以 1, $b_1, b_2, \cdots, b_n$ 为系数且其根为 $x_1^2, x_2^2, \cdots, x_n^2$ 的多项式后，我们同样可以得到以多项式

$$y^n+b_1y^{n-1}+b_2y^{n-2}+\cdots+b_n$$

的根的平方(即 $x_1^4, x_2^4, \cdots, x_n^4$) 为根的多项式，它的系数是 $1, c_1, c_2, \cdots, c_n$. 类似地可以进一步得到以 $1, d_1, d_2, \cdots, d_n$ 为系数而其根为 $x_1^8, x_2^8, \cdots, x_n^8$ 的多项式，然后得到以 $x_1^{16}, x_2^{16}, \cdots, x_n^{16}$ 为根的多项式等等.

为了简单起见，我们只就方程的根全是实数且其绝对值都不相同的情形来考察罗巴切夫斯基方法的基本想法. 设

$$|x_1|>|x_2|>\cdots>|x_n|,$$

即 x_1 是绝对值最大的根，x_2 的绝对值其次，等等，设 N 是充分大的正整数，而多项式

$$X^n+A_1X^{n-1}+A_2X^{n-2}+\cdots+A_n$$

的根是给定的多项式的根 $x_1, x_2, \cdots, x_n$ 的 N 次方，即

$$-A_1=x_1^N+x_2^N+\cdots+x_n^N,$$

$$A_2=x_1^Nx_2^N+x_2^Nx_3^N+\cdots+x_{n-1}^Nx_n^N,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\pm A_n=x_1^Nx_2^N\cdots x_n^N.$$

于是当 N 充分大时，在数串 $|x_1^N|, |x_2^N|, \cdots, |x_n^N|$ 中，每一个后面的数较之前面的数是这样的小，以致在考虑 $A_1, A_2, \cdots, A_n$ 的式子时，可以只留下第一项，而其余各项的和按比值来讲可以忽略不计，于是我们得到近似公式

$$x_1^N\approx -A_1,\quad x_1^Nx_2^N\approx A_2,$$

$$x_1^Nx_2^Nx_3^N\approx -A_3, \cdots, x_1^Nx_2^Nx_3^N\cdots x_n^N\approx \pm A_n.$$

将它们成对地逐项相除后开 n 次方，于是对于 x_k 就得到下列诸公式:

$$x_1=\sqrt[N]{-A_1},\quad x_2=\sqrt[N]{-\frac{A_2}{A_1}},\quad x_3=\sqrt[N]{-\frac{A_3}{A_2}},\ \cdots,$$

$$x_n=\sqrt[N]{-\frac{A_n}{A_{n-1}}}.$$

能够证明：在计算了足够多次以后，得到的多项式的系数在依次加上符号 $+, -, +, -, \cdots$ 后，就足够地近似于前一多项式的对应

的系数的平方.

罗巴切夫斯基方法的详细阐述可以在克雷洛夫院士的著名著作«近似计算讲义»中找到.

文　献

通俗小册子

А. Г. Курош, Алгебраические уравнения произвольных степеней. Гостехиздат, 1951.

А. И. Маркушевич, Комплексные числа и конформные отображения. Гостехиздат, 1954.

И. Р. Шафаревич, О решении уравнений высших степеней (Метод Штурма). Гостехиздат, 1954.

大学教程

А. Г. 库洛什,高等代数教程,高等教育出版社, 1953 年版。

Л. Я. 奥库涅夫,高等代数,高等教育出版社, 1953 年版。

А. К. Сушкевич, Основы высшей алгебры. Гонти, 1937.

(这些书中的第一本较着重于代数的抽象方面,而其余两本则较多地注意具体的内容.)

严士健 译

秦元勋 校

第二卷

数学名著译丛

数 学

它的内容，方法和意义

〔俄〕A. D. 亚历山大洛夫 等 著

科学出版社
www.sciencep.com

内 容 简 介

本书是苏联数学界普及数学知识的一部名著．全书共二十章，分三卷出版．书中不仅介绍了近代数学的各个分支，还介绍了数学的哲学、历史发展及其在物理学和工程技术方面的应用．本书每章都由苏联第一流的学者撰著，有较高的学术水平．同时，本书写得深入浅出，只要具备高中数学知识，就可以理解．在努力实现四个现代化的今天，这部书必将帮助我国广大读者扩大眼界，找到解决问题的数学工具．本卷共十章，包括微分方程、变分法、复变函数、数论、概率论、函数逼近、计算方法和电子计算机等方面的内容．

本书可供大学数学系师生、中学教师、具有高中以上文化程度的学生、有关方面的科技人员及数学爱好者阅读．

А. Д. Александров

МАТЕМАТИКА, ЕЁ СОДЕРЖАНИЕ, МЕТОДЫ И ЗНАЧЕНИЕ/ТОМ 2/

Издательство Академии Наук СССР, 1956

数 学

——它的内容、方法和意义——

第 二 卷

〔苏〕А. Д. 亚历山大洛夫 等 著

秦元勋 王光寅 裘光明 胡祖炽 越民义 许孔时 梁文祺 余家荣 储锺武 于桂芝 译

周毓麟 孙以丰 冯 康 许孔时 越民义 王寿仁 秦元勋 校

*

科学出版社 出版

北京朝阳门内大街 137 号

中国科学院印刷厂 印刷

新华书店北京发行所发行 各地新华书店经售

*

1959 年 4 月第 一 版　　开本：850×1168 1/32

1984 年 7 月第五次印刷　　印张：12 7/8

印数：17,571—43,670　　字数：340,000

统一书号：13031・2494

本社书号：3427・13—1

定价：1.90 元

第 一 卷

第 一 章　数学概观（А. Д. 亚历山大洛夫著）
第 二 章　数学分析（М. А. 拉夫伦捷夫 / С. М. 尼阔尔斯基 合著）
第 三 章　解析几何（Б. Н. 狄隆涅著）
第 四 章　代　　数——代数方程的理论（Б. Н. 狄隆涅著）

第 三 卷

第十五章　实变函数（С. Б. 斯捷奇金著）
第十六章　线性代数（Д. К. 法德杰也夫著）
第十七章　抽象空间（А. Д. 亚历山大洛夫著）
第十八章　拓扑学（П. С. 亚历山大洛夫著）
第十九章　泛函分析（И. М. 盖尔芳特著）
第二十章　群及其他代数系统（А. И. 马尔采夫著）

目　　录

第　二　卷

第五章　常微分方程(И. Г. 彼得罗夫斯基著) ……………………1

§1. 绪论……………………………………………………………1

§2. 常系数线性微分方程 ……………………………………12

§3. 微分方程的解及应注意的几个方面 ……………………20

§4. 微分方程积分问题的几何解释. 问题的推广 …………22

§5. 微分方程解的存在性与唯一性方程的近似解 …………25

§6. 奇点 ………………………………………………………32

§7. 常微分方程定性理论 ……………………………………37

第六章　偏微分方程(С. Л. 索伯列夫著)……………………48

§1. 绪论 ………………………………………………………48

§2. 最简单的数学物理方程 …………………………………50

§3. 始值条件和边值条件. 解的唯一性 ……………………59

§4. 波的传播 …………………………………………………69

§5. 解法 ………………………………………………………71

§6. 广义解(О. А. 拉窦席斯加娅著)………………………91

第七章　曲线和曲面(А. Д. 亚历山大洛夫著)………………97

§1. 关于曲线和曲面理论的对象和方法的概念 ……………97

§2. 曲线理论…………………………………………………101

§3. 曲面理论的基本概念……………………………………116

§4. 内蕴几何和曲面的弯曲变形……………………………131

§5. 曲线和曲面理论中的新方向……………………………149

第八章　变分法(В. И. 克雷洛夫著)…………………………159

§1. 绪论………………………………………………………159

§2. 变分法的微分方程………………………………………164

§3. 变分法问题的近似解法…………………………………175

第九章　复变函数(М. В. 凯尔迪什著) ……………………178

§1. 复数和复变函数…………………………………………178

§2. 复变函数与数学物理问题的关系……………………………191
§3. 复变函数与几何的关系……………………………………201
§4. 线积分. 柯西公式及其推论………………………………212
§5. 唯一性和解析拓展…………………………………………224
§6. 结论……………………………………………………………231
第十章 素数(К. К. 马尔德尔扎尼吉维里著) ………………233
§1. 数论研究什么和如何研究数论…………………………233
§2. 如何研究与素数有关的问题………………………………238
§3. 关于车比雪夫方法…………………………………………245
§4. 维诺格拉朵夫方法…………………………………………251
§5. 整数分解为二平方之和. 整复数(А. Г. 波斯特尼可夫著)……259
第十一章 概率论(А. Н. 柯尔莫果洛夫著) ……………………263
§1. 概率规律性…………………………………………………263
§2. 初等概率论的公理与基本公式……………………………265
§3. 大数定律与极限定理………………………………………272
§4. 关于概率论基本概念的补充说明…………………………282
§5. 因果过程与随机过程………………………………………288
§6. 马尔科夫型的随机过程……………………………………294
第十二章 函数逼近法(С. М. 尼阔尔斯基著) ………………299
§1. 绪论……………………………………………………………299
§2. 插值多项式…………………………………………………303
§3. 定积分的逼近………………………………………………310
§4. 车比雪夫最好一致逼近的观念……………………………316
§5. 与零偏差最小的车比雪夫多项式…………………………319
§6. 魏尔斯特拉斯定理. 函数的最好逼近与它的微分性质…………322
§7. 傅里叶级数…………………………………………………325
§8. 在平均平方意义下的逼近…………………………………332
第十三章 近似方法与计算技术(В. И. 克雷洛夫著) ………338
§1. 近似及数值的方法…………………………………………338
§2. 最简单的计算辅助工具……………………………………353
第十四章 电子计算机(С. А. 勒贝杰夫著) ……………………365
§1. 电子计算机的功用和基本工作原理………………………365
§2. 在快速电子计算机中的程序设计和代码的编制……………371
§3. 快速计算机部件的技术原理在电子计算机上执行运算的次序…383
§4. 电子计算机的发展和使用的远景(Л. В. 康托洛维奇著)………398

第五章 常微分方程

§1. 绪 论

微分方程的例子 在前几章中我们所遇到的方程主要是求某几个量的数值. 例如在求函数的极大值和极小值时，我们需要解方程去求函数的变化率为零的点；又如在第四章(第一卷)中研究求多项式的根的问题等等.所有这些情形都是在求个别的数值.但是在数学的应用中，时常遇到性质上完全新的问题，在这类问题中,函数本身是未知的,一些变数对另一些变数的依存规律本身是未知的. 例如研究物体的冷却过程，我们要确定它的温度如何随时间而变化;在决定行星或星球的运动时,我们要决定它的坐标与时间的依存关系等等.

对于所要找的未知函数,我们时常可以作出它的方程,这类方程称为泛函方程，一般说，泛函方程的性质可以是各种各样的(研究稳函数的问题可以说是我们已经遇到过的最简单最原始的泛函方程).

在第五、六及八章中,将研究未知函数的寻求问题. 在这一章及下章中要研究未知函数的方程中最重要的方程，即所谓微分方程. 从这个名称就可以知道,这个方程中不只出现函数本身,而且还出现它的某些阶的微商.

下面的等式可以作为微分方程的例子:

$$\frac{dx}{dt}+P(t)x=Q(t),\quad \frac{d^2x}{dt^2}+m^2x=A\sin\omega t$$

$$\frac{d^2x}{dt^2}=tx,\quad \frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2},\quad \frac{\partial^2 u}{\partial t^2}=\frac{\partial^2 u}{\partial x^2},$$

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0. \tag{1}$$

前面三个方程用 x 记作未知函数，而 t 记作独立变数；后面三个方程用 u 记作未知函数，这些函数依赖于两个变数 x 与 t 或 x 与 y.

对于数学，特别是对于数学的应用，微分方程所具有的重大意义主要是在于：很多物理问题与技术问题的研究可以化归为这类方程的求解问题.

电子计算机及无线电装置的计算，弹道的计算，飞机在飞行中稳定性的研究，或者化学反应过程的稳定性的研究，所有这些都可以化为微分方程求解的问题来进行.

某些物理现象所服从的物理规律可用微分方程表现出来，而这些微分方程则是这些规律的准确的量的（数值的）表示的工具，这种情形是经常最易遇到的. 读者在下一章中将会看到，例如用微分方程之形式表示出质量守恒定律与热能守恒定律. 借助于微分方程，牛顿所发现的力学规律可以用来研究所有力学系统的运动.

我们用简单的例子来说明这一点. 设所研究的质点具有质量 m，在 Ox 轴上运动，在时间 t 它的坐标用 x 来表示.

当质点运动时，它的坐标 x 也在随时间而变化，而要知道质点的运动也即是要知道 x 对于时间 t 的函数依存关系. 设运动是在力 F 作用下进行的，又设力 F 是与用坐标 x 所定义的质点的位置，与运动的速度 $v=\frac{dx}{dt}$ 及与时间 t 有关的，即是 $F=F\left(x,\frac{dx}{dt},t\right)$. 依照力学规律，作用于质点上的力 F 将引起运动的加速度 $w=\frac{d^2x}{dt^2}$，使得质点的质量 m 乘这个加速度确切地等于作用力的大小，也即是，在运动的任何时间，等式

$$m\frac{d^2x}{dt^2}=F\left(x\cdot\frac{dx}{dt},\ t\right) \tag{2}$$

是应该满足的.

这便是描述质点运动过程的函数 $x(t)$ 所应满足的微分方程式. 它是上述力学规律的简单的描述. 它的意义在于可以把决定

质点运动的力学问题化为微分方程求解的数学问题.

下面读者可以找到其他的例子，说明各种不同的物理过程的研究如何可以化为微分方程的研究.

微分方程理论在十七世纪的末年开始发展起来，差不多是与微分及积分的计算同时产生的. 现在微分方程已经成为研究自然现象的强有力的工具. 在力学、天文学、物理学及技术科学中借助于微分方程已经取得了巨大的成就. 牛顿研究天体运动的微分方程,从理论上得到行星运动规律,而这些规律原来只是由凯普勒实验中得到的. 勒未累在1846年预言海王星的存在,并在这个微分方程数值分析的基础上,决定海王星在天空中的位置.

为了说明微分方程理论的一般特点，首先要指出，一般说来，每一个微分方程不只有一个解而是有无限多个解，而是有无限多个函数满足这个方程. 例如,在前面所说的力学例子中,不论开始时质点所在的位置如何以及开始时的速度如何，只要在同样的函数 $F\left(x, \frac{dx}{dt}, t\right)$所表示的力作用下,所有的运动都必须满足前述的质点运动的方程,对应于每个运动有一个 x 对时间 t 的依存关系. 因为在力 F 作用下的运动可以有无限多个，微分方程 (2) 也将有无限多个解.

一般说来，每一个微分方程确定一整族的满足它的函数. 微分方程理论的基本问题是研究满足这个微分方程的函数. 微分方程理论使得有可能充分全面地表达出满足方程的所有函数的性质. 这在自然科学的应用上是特别重要的. 此外,如果需要计算,微分方程理论应保证有办法算出函数的数值. 如何做到这些，下面将会谈到.

如果未知函数只与一个变数有关，这种微分方程称为**常微分方程**. 而当未知函数与几个变数有关，方程中又出现未知函数对几个变数的微商,这种微分方程称为**偏微分方程**. 在(1)中前三个方程是常微分方程,后三个是偏微分方程.

偏微分方程理论具有本质上与常微分方程理论不同的许多固

有特点．有关偏微分方程的基本概念将在下一章中述及；本章中我们只谈到常微分方程．

现在看看下面几个例子．

例 1 镭衰变的规律是：衰变速度与镭所存余的量成比例．设已知在某一时间 $t=t_0$ 有 R_0 克的镭．要确定在任何时间 t 镭的量．

设 $R(t)$是在时间 t 尚未衰变的镭的量，衰变速度用量 $-\frac{dR}{dt}$ 来计量，因为速度与 R 成比例，我们得到

$$-\frac{dR}{dt}=kR, \tag{3}$$

式中 k 是一个常量．

要解决我们的问题便要由方程(3)决定函数 R，为此，注意到 $R(t)$的倒函数满足方程

$$-\frac{dt}{dR}=\frac{1}{kR}, \tag{4}$$

因为$\frac{dt}{dR}=\frac{1}{\frac{dR}{dt}}$．由积分计算知道具有形式

$$t=-\frac{1}{k}\ln R+C$$

的任意函数都满足方程(4)，式中 C 为任意常数．由此关系，我们定出 t 的函数 R．我们有

$$R=e^{-kt+kC}=C_1e^{-kt}. \tag{5}$$

由方程(3)的所有的解(5)中要提出在 $t=t_0$ 时取 R_0 值的那个解，如果将 $C_1=R_0e^{kt_0}$ 代入(5)我们就得到了这个解．

从数学的观点看来，方程(3)是函数 R 变化的非常简单的规律的描述，也就是说，函数 R 的减少速度 $-\frac{dR}{dt}$ 与函数 R 本身的值成比例．不仅在放射性衰变现象中，而且在许多其他物理现象中，满足这种函数变化的规律．

例如，我们在物体冷却的研究中也遇到函数变化的同样的规

律，此时物体中的热量的减少是和物体的温度与其周围介质的温度之差成比例，在许多其他物理过程的研究中，也有这类规律．因此，虽然我们由镭的衰变引出方程(3)，但方程(3)应用的范围比镭的衰变这个特殊问题更有无比的广泛性．

例 2 设具有质量 m 的质点沿水平轴 Ox 运动于有阻力的介质中，例如在液体或气体中，受到依虎克定律作用的两个弹簧的弹性力的影响(图 1)．这个定律即是弹力向平衡位置起作用并与离平衡位置之偏差成比例．设平衡位置在点 $x=0$．则弹性力等于 $-bx(b>0)$．

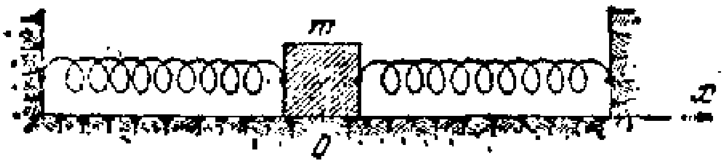


图 1

设介质阻力与运动之速度成比例，即等于 $-a\frac{dx}{dt}$，其中 $a>0$，又负号表示介质阻力的方向与运动速度的方向相反．当速度小时，关于介质阻力的这个假设与实验很符合．

基于牛顿定律质点的质量乘其加速度等于加于此质点的力的和，即是

$$m\frac{d^2x}{dt^2}=-bx-a\frac{dx}{dt}. \tag{6}$$

因此，在任何时间 t 表示质点运动位置的函数 $x(t)$ 满足微分方程(6)．我们将在后面的一节中研究这个方程．

如果除了上述各力外，还对质点施以外力 F，则运动的微分方程(6)将采取以下形式：

$$m\frac{d^2x}{dt^2}=-bx-a\frac{dx}{dt}+F. \tag{6'}$$

例 3 数学摆即是一个质量 m 的质点系于一个长度为 l 的线上(图 2)．我们假设在任何时候这个摆都在同一平面上——即图画所在的平面上．作用于质点的重力 mg 是将摆拉回平衡位置

OA 的力．摆在任何时间 t 的位置可以用它与垂线 OA 相差的角 φ 来确定．反时钟运动的方向作为 φ 角计算的正方向，设质点由平衡位置 A 所经的道路为弧 $AA'=l\varphi$．速度 v 将是沿圆的切线方向并将取以下的数值

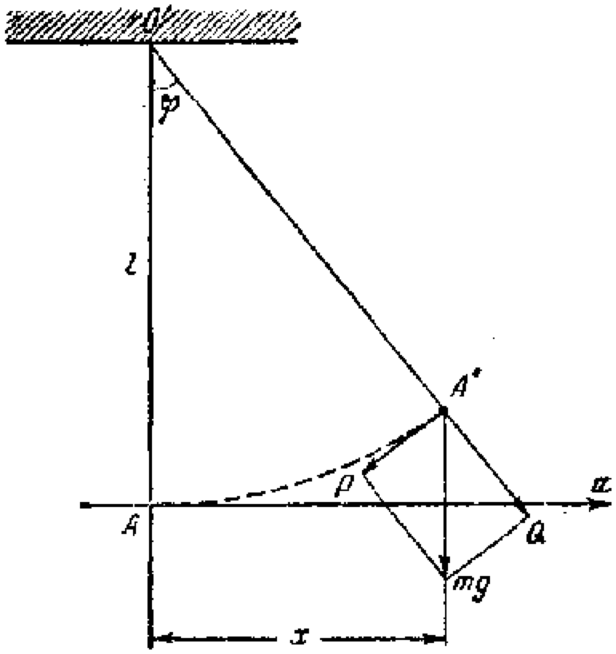


图 2

$$v=l\frac{d\varphi}{dt}.$$

为了作出运动的方程，把重力 mg 分为两个分量 Q 与 P，第一个分量沿着 OA' 半径方向，第二个分量沿着圆周的切线方向．分量 Q 不会引起速度 v 的数值的改变，因为它与相反方向的拉力 OA' 相抵消．只有分量 P 引起速度 v 的数值的变化．它总是向着平衡位置 A 起作用，即是当角 φ 为正时，向减小 φ 的方向，当角 φ 为负时，向增加 φ 的方向．P 的数值等于 $-mg\sin\varphi$，因此摆的运动方程是

$$m\frac{dv}{dt}=-mg\sin\varphi$$

或

$$\frac{d^2\varphi}{dt^2}=-\frac{g}{l}\sin\varphi. \tag{7}$$

有趣的是，这个方程的解不能用有限形式表示为初等函数．甚至就数学摆的振动这类简单的物理过程来说，要想用初等函数来给出它的准确表达式，也已经表示出初等函数所包含的范围是很贫乏的．后面我们要看到，用初等函数可求解的微分方程是不多的．常常会有这样的情形，在物理或力学中遇到的一些微分方程的研究迫使我们引入新的函数，对这些函数进行研究，从而将解决应用问题时所用的函数的范围加以扩大．

现在我们只研究摆的微小振动，即当忽视微小的误差后，把 AA' 弧看作它在 Ox 轴上的投影 x，把 φ 代替 $\sin\varphi$ 的情况．则 $\varphi\approx\sin\varphi=\frac{x}{l}$ 及摆的运动方程取更简单的形式

$$\frac{d^2x}{dt^2}=-\frac{g}{l}x. \tag{8}$$

下面我们要说明，这个方程可用三角函数来解，由此可能相当准确地描述摆的"微小振动"．

例 4 赫尔姆霍尔茨的声学共振器(图 3)是由容积为 v 的充满空气的容器 V 和圆柱形的颈部 F 所组成．近似地可将颈部 F 中的空气的质量看作具有质量 m 的柱子，

$$m=\rho sl, \tag{9}$$

式中 ρ 为空气密度，s 为颈部的横截面，l 为其长度．

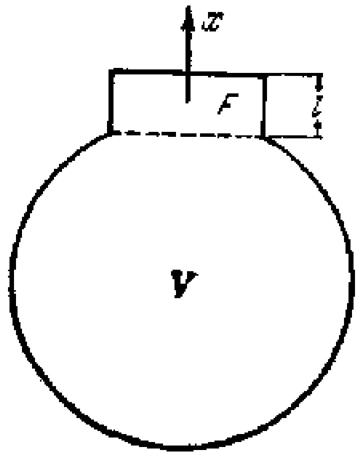


图 3

设空气从平衡位置移动的大小为 x，则具有容积 v 的容器的空气压力由初始压力变化一些，以 Δp 记此变化．

假设压力 p 与体积 v 由绝热变化规律 $pv^k=C$ 相联系．则如果略去高阶的小量，我们得到

$$\Delta p\cdot v^k+pkv^{k-1}\cdot\Delta v=0$$

及

$$\Delta p=-kp\frac{\Delta v}{v}=-\frac{kps}{v}x. \tag{10}$$

(在我们这种情形 $\Delta v=sx$)．在颈部中空气的运动方程可以写成

$$m\frac{d^2x}{dt^2}=\Delta p\cdot s, \tag{11}$$

式中 $\Delta p\cdot s$ 是在容量中的气体加于颈部中的空气柱的压力．由关系(10)及(11)我们得到

$$\rho l\frac{d^2x}{dt^2}=-\frac{kps}{v}x, \tag{12}$$

式中 ρ, p, v, l, k, s 都是常数．

例 5 在最简单的振动回路中的电振荡的研究也化为形式(6)的方程. 这个回路的图式如图 4 所表示. 这里左边表示电容为 C 的电容器, 它的极板和自感 L 及电阻 R 串联成闭合回路. 设在某一时刻电容器得到一个电位差, 此后其电源即切断. 如果回路中不存在自感, 则电流就流动到电容器内不存在电位差时为止.

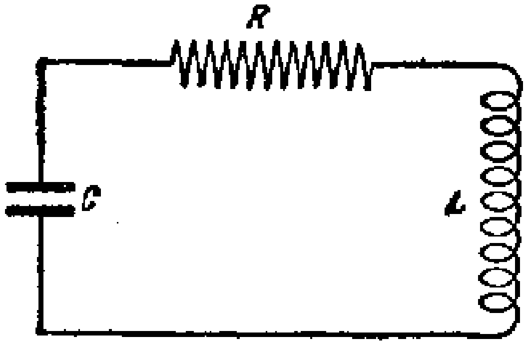


图 4

当有自感存在时, 则过程便完全两样、在回路中产生电振荡. 为要引出这个振荡的规律, 以 $v(t)$ 或简单地用 v 表示在时间 t 电容器极板上的电位差, 以 $I(t)$ 表示在时间 t 的电流强度, 以 R 表示电阻. 由熟知的物理规律, 在任何时间 $I(t)R$ 均等于全部电动势, 而后者包括电容器极板上电位差所产生的电动势 $-v(t)$ 以及自感电动势 $-L\frac{dI}{dt}$. 由此有

$$IR=-v-L\frac{dI}{dt}. \tag{13}$$

以 $Q(t)$ 记在时间 t 电容器上的电量. 则在回路上每时的电流强度等于 $\frac{dQ}{dt}$. 电容器极板上的电位差 $v(t)$ 等于 $\frac{Q(t)}{C}$. 因此, $I=\frac{dQ}{dt}=C\frac{dv}{dt}$, 等式(13)则改写为

$$LC\frac{d^2v}{dt^2}+RC\frac{dv}{dt}+v=0. \tag{14}$$

例 6 图 5 中所表示的为电磁振荡的真空管振荡器的线路图, 由电容 C, 电阻 R 及自感 L 所组成的振荡回路作为基本振荡系统. 线圈 L' 及在图 5 中心所示的真空管组成所谓的反馈. 电池 B 是和 L, R, C 相联的能源, k 是真空管的阴极, A 为阳极, S 为栅极. 在这种线路图中, 在 L, R, C 回路中产生"自振". 在所有的实际的系统中, 在振动的过程, 能量必变成热或用其他形式传

到周围的物体上去．因此，为要保持振动的驻定状况，为了保持振幅，每一个实际的振动系统必须从外面取得能量．自振与其他振动过程的区别在于，为了保持这个系统的驻定振动状况，外扰不必是周期的．自振系统的装置要有个经常的能源，在我们的例子中电池 B 保持驻定振动状况．钟表、电铃、音乐家手中的弦与弓、人的声带及其他例子都是自振系统．

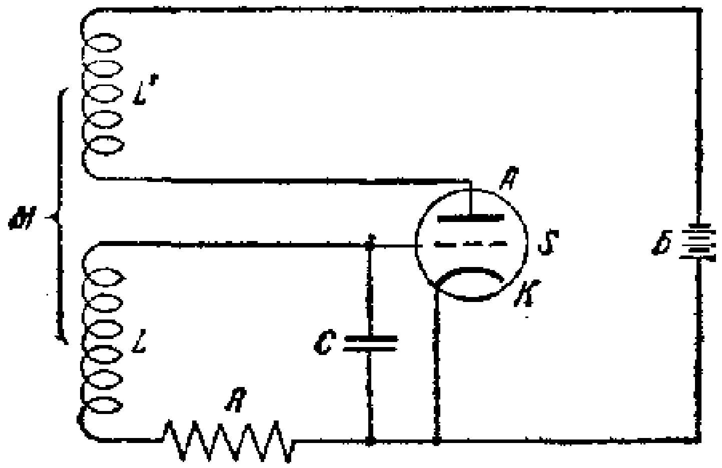


图 5

在 L, R, C 振动回路中的电流强度 $I(t)$ 满足方程

$$L\frac{dI}{dt}+RI+v=M\frac{dI_a}{dt}, \tag{15}$$

式中 $v=v(t)$ 是在时间 t 电容器极板上的电位差，$I_a(t)$ 是通过线圈 L' 之阳极电流；M 是线圈 L 与 L' 之间的互感系数．方程(15)比方程(13)只多了一项 $M\frac{dI_a}{dt}$．

可设阳极电流 $I_a(t)$ 只和真空管的栅极 S 与阴极之间的电位差有关，即是忽略了阳极的反作用；在这个假定下，这个电位差等于电容器 C 极板上的电位差 $v(t)$．在图 6 中表示了 I_a 与 v 的函数相关的特性曲线通常用三次抛物线来表示这个曲线，近似地可写成

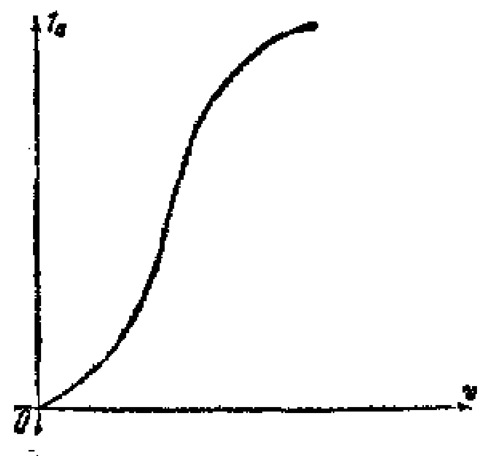


图 6

$$I_a = a_1 v + a_2 v^2 + a_3 v^3.$$

将此式代入方程(15)的右边，又利用

$$\frac{dv}{dt} = I,$$

我们得到 v 的方程

$$L\frac{d^2v}{dt^2} + [R - M(a_1 + 2a_2 v + 3a_3 v^2)]\frac{dv}{dt} + v = 0. \tag{16}$$

在这些例子中，对于已给的物理过程，寻找其示性的物理量的问题均化为求常微分方程解的问题.

微分方程论的问题 现在来给出确切的定义. 在变数 x 与值

$$y(x),\ y'(x) = \frac{dy}{dx},\ y''(x) = \frac{d^2y}{dx^2},\ \cdots,\ y^{(n)}(x) = \frac{d^ny}{dx^n}$$

之间的形如

$$F[x,\ y(x),\ y'(x),\ y''(x),\ \cdots,\ y^{(n)}(x)] = 0 \tag{17}$$

的关系式，称为**一个未知函数 y 的 n 阶的常微分方程**. 微分方程中，出现的未知函数的最高次微商，称为**微分方程的阶**. 因此，在例 1 中所得到的微分方程是一阶，在例 2、3、4、5、6 中的是二阶.

如果将函数 $\varphi(x)$ 代替(17)中的 y，$\varphi'(x)$ 代替 y'，…，$\varphi^{(n)}(x)$ 代替 $y^{(n)}$，使得(17)变成恒等式，则函数 $\varphi(x)$ 称为**微分方程(17)的解.**

物理与技术的问题时常化为常微分方程组，其中含有几个依赖于同一个独立变数的未知函数以及它们的微商.

为了说得更具体起见，以下主要只谈一个未知函数的不超过二阶的常微分方程，用这个例子来说明，所有的常微分方程及未知函数的个数，等于方程的个数的方程组的基本性质.

上面我们已经说过，每一个常微分方程不是只有一个解，而是有无限多个解. 现在回到这个问题上来，首先用前面 2 到 6 的例子来说明一些明显的看法. 在每个例子中所研究的与物理系统对应的方程，只是由这个系统的结构所完全确定. 但在这些系统的每一个当中，可以发生不同的过程. 例如，十分显然，依照方程(8)

运动的摆可以作非常不相同的振动. 摆的每一个振动与方程(8)的一个解相对应,因此必有无限多个这样的解. 可以验证,形如

$$x=C_1\cos\sqrt{\frac{g}{l}}\,t+C_2\sin\sqrt{\frac{g}{l}}\,t \tag{18}$$

的任何函数都满足方程(8),其中 C_1 与 C_2 是任意常数.

物理上也是很明显的,只有当我们在某一个时间 t_0 给定 x 等于 x_0 的(初始)值(即物质质点与平衡位置最初的偏离)以及运动的初始速度 $x'_0=\left.\frac{dx}{dt}\right|_{t=t_0}$ 时,摆的运动才是完全确定的. 用这些初始条件可以决定公式(18)中的常数 C_1 与 C_2.

我们在其他例中所得的微分方程完全同样会有无穷多个解.

关于具有一个未知函数的 n 阶的微分方程(17),一般在很广泛的假定下可以证明有无数多个解. 更确切地说:如果对于独立变数的某个"初始值",我们给定未知函数及其直到 $n-1$ 次微商的"初始值",则可找到方程(17)的满足这些"给定初始值"的解. 也可以证明,用这些初始条件完全决定了解,只存在一个满足上面已给初始条件的解. 关于这些我们后面将要更详细地讲到. 现在我们的目的是要指明,函数及其 $n-1$ 个导数的初始值是可以任意给的. 我们可以任意选择,作为所求解的"初始状态"的 n 个量.

设有可能性,我们把某个 n 阶微分方程的所有的解用一个公式表示出来,则这种公式中必定包含刚好 n 个独立任意常数,由这些数的选取我们可以满足 n 个初始条件. n 阶微分方程的这种包含有 n 个独立常数的解普通常称为这个**微分方程的通解**. 例如方程(8)的通解为公式(18),其中有两个任意常数,由公式(5)得到方程(3)的通解.

现在我们来看微分方程理论所遇到的问题的最一般的特点. 它们是各种各样的. 我们来谈其中最主要的.

设对微分方程再加上初始值,则微分方程的解便完全确定了. 要求明显地表出解的公式,是微分方程理论的首要问题之一. 只有在最简单的情形才能得到这种公式,但如果找到了这种公式,

则对解的计算与研究都有很大的帮助.

理论应当对于形成解的特性的概念提供可能性：解是否是单调的或是振动的？是否是周期的或趋近于周期的等等.

很明显，当我们将未知函数及其微商的初始值加以变动时(即被研究的系统的初始状态被改变)则解本身也将改变(即将换为另一个过程). 微分方程理论应当保证我们有可能讨论这种变化是怎样的. 特别是，对于初始值的微小变化，解本身的变化会不会是很小的，因此，解在这种情形下会不会是稳定的，或者初始值的微小变化可能引起解本身的强烈的变化，即它是不稳定的.

我们也应当不只是对于方程的每个单独的解，而是对于它的所有的解的全体得出定性的以及如果可能时得出定量的分布图形. 在设计时常常发生选取参数的问题，这些参数是仪器或机械的表征，选出它以保证工作的良好进行. 仪器的参数是在描述这种仪器工作的微分方程中以某种形式的量出现. 理论应当帮助我们说明，如果我们变化方程本身(变化其中出现的仪器参数)，(仪器工作的)解将发生些什么影响.

最后，当需要进行计算时，即要用数值计算来求微分方程的解时，理论应当提供给工程师及物理学家以可能更经济的与更快速的解的计算方法.

§2. 常系数线性微分方程

常微分方程中有一些重要的类型，其通解可以用简单的已经很好的研究过的函数表示出来. 这些类型中的一类是对于未知函数及其微商是线性的(简称为线性的)常系数微分方程. 例如，微分方程(3)，(6)，(8)，(14)便是这种类型. 如果方程中不具有与未知函数无关的项，则这个线性方程称为**齐次的**，如具有这种项，则称为**非齐次的**.

二阶常系数齐次线性方程 这种方程具有下面的形式

$$m\frac{d^2x}{dt^2}+a\frac{dx}{dt}+bx=0, \tag{6}$$

式中 m, a, b 是常数. 不妨设 m 为正的; 这并不失其普遍性, 因为我们已假设 $m\neq0$, 如果 m 为负的, 我们只要将所有的系数都换一个负号便可以达到这种假定.

对这个方程来找具有指数函数 $e^{\lambda t}$ 形式的解, 要找常数 λ 使得函数 $x=e^{\lambda t}$ 满足方程. 将 $x=e^{\lambda t}$, $\frac{dx}{dt}=\lambda e^{\lambda t}$, 及 $\frac{d^2x}{dt^2}=\lambda^2e^{\lambda t}$ 代入方程(6)的左方, 得到

$$e^{\lambda t}(m\lambda^2+a\lambda+b).$$

因此, 要 $x(t)=e^{\lambda t}$ 是方程(6)的解, 充分而且必要的条件是

$$m\lambda^2+a\lambda+b=0. \tag{19}$$

如果 λ_1 与 λ_2 是方程(19)的两实根, 则容易验证, 所有形式为

$$x=C_1e^{\lambda_1 t}+C_2e^{\lambda_2 t} \tag{20}$$

的函数都是方程(6)的解, 其中 C_1 与 C_2 是任意常数.

下面我们来证明, 如果方程(19)有不同的实根, 则公式(20)表示方程(6)的所有的解.

我们注意到方程(6)的下面的重要的性质:

1. 方程(6)的两个解的和也是这个方程的解.

2. 用常数乘方程(6)的解, 得到的也是一个解.

如果 λ_1 是方程(19)的重根的情形, 即是 $m\lambda_1^2+a\lambda_1+b=0$ 与 $2m\lambda_1+a=0$[1], 则函数 $te^{\lambda_1 t}$ 也将是方程(6)的解, 因为, 将这个函数及其导函数代入方程(6)的左边, 我们得到

$$te^{\lambda_1 t}(m\lambda_1^2+a\lambda_1+b)+e^{\lambda_1 t}(2m\lambda_1+a),$$

由前述的两个等式, 这里便恒等于零了.

在这种情形下, 方程(6)的通解为

$$x=C_1e^{\lambda_1 t}+C_2te^{\lambda_1 t}. \tag{21}$$

现在设方程(19)有复数根. 因为 m, a, b 均为实数, 这两个

---

1) 二次方程(19)的根 λ_1 与 λ_2 的和为 $\lambda_1+\lambda_2=-\frac{a}{m}$, 如果两根相等, 即 $\lambda_1=\lambda_2$, 则由此可导出第二个等式.

复根便是共轭的．设 $\lambda=\alpha\pm i\beta$．方程

$$m(\alpha+i\beta)^2+a(\alpha+i\beta)+b=0$$

等价于两个等式

$$m\alpha^2-m\beta^2+a\alpha+b=0 \text{ 及 } 2m\alpha\beta+a\beta=0. \tag{22}$$

在这种情形下，容易验证函数 $x=e^{\alpha t}\cos\beta t$ 及 $x=e^{\alpha t}\sin\beta t$ 是方程(6)的解．实际上，例如将函数 $x(t)e^{\alpha t}\cos\beta t$ 及其微商代入方程(6)的左边，我们得到

$$e^{\alpha t}\cos\beta t(m\alpha^2-m\beta^2+a\alpha+b)-e^{\alpha t}\sin\beta t(2m\alpha\beta+a\beta),$$

而由等式(22)这个表示便恒等于零了．

如果方程(19)有复根，方程(6)的通解便可写成

$$x=C_1e^{\alpha t}\sin\beta t+C_2e^{\alpha t}\cos\beta t, \tag{23}$$

式中 C_1 与 C_2 是任意常数．

方程(19)称为**示性方程**，知道了方程(19)的根，我们可以写出方程(6)的通解．

我们注意到，对于常系数 n 阶线性齐次方程

$$a_n\frac{d^nx}{dt^n}+a_{n-1}\frac{d^{n-1}x}{dt^{n-1}}+\cdots+a_1\frac{dx}{dt}+a_0x=0,$$

如果已知道代数方程，也称为**示性方程**

$$a_n\lambda^n+a_{n-1}\lambda^{n-1}+\cdots+a_1\lambda+a_0=0$$

的根时，则微分方程的通解可以类似地用多项式，指数函数及三角函数表示出来．因此，常系数线性齐次方程的积分问题归结为代数问题．

现在我们来证明公式(20)，(21)，(23)实际上给出了方程(6)的所有的解．注意，这些公式中的 C_1 与 C_2 经常可以这样选取，使得函数 $x(t)$ 满足任何初始条件

$$x(t_0)=x_0,\ x'(t_0)=x'_0.$$

由公式(20)可以由方程组

$$x_0=C_1e^{\lambda_1 t}+C_2e^{\lambda_2 t_0},$$

$$x'_0=\lambda_1C_1e^{\lambda_1 t_0}+\lambda_2C_2e^{\lambda_2 t_0}$$

来决定 C_1 与 C_2，对公式(21)与(23)也是类似的。因此，如果方程(6)有这样的解，它不会在上述的公式中，则对于同样的初始条件，方程(6)有两个不同的解。这两个不同的解的差可用 $x_1(t)$ 表示，则 $x_1(t)$ 将满足方程(6)，并且满足初始条件 $x_1(t_0)=0$, $x_1'(t_0)=0$，而 $x_1(t)$ 不恒等于零。现在我们来证明，满足初始条件为零的方程(6)的解只有 $x_1(t)\equiv 0$。首先假设 $m>0$, $a>0$, $b>0$。用 $2\dfrac{dx_1}{dt}$ 乘等式

$$m\frac{d^2x_1}{dt^2}+a\frac{dx_1}{dt}+bx_1=0 \tag{24}$$

的左右两方。因 $2\dfrac{dx_1}{dt}\cdot\dfrac{d^2x_1}{dt^2}=\dfrac{d}{dt}\left(\dfrac{dx_1}{dt}\right)^2$ 及 $2x_1(t)\dfrac{dx_1}{dt}=\dfrac{d}{dt}(x_1^2)$，故等式(24)可以写成

$$\frac{d}{dt}\left[m\left(\frac{dx_1}{dt}\right)^2\right]+2a\left(\frac{dx_1}{dt}\right)^2+b\frac{d}{dt}(x_1^2)=0.$$

将这个等式从 t_0 积分到 t，我们得到

$$m\left(\frac{dx_1}{dt}\right)^2+2a\int_{t_0}^{t}\left(\frac{dx_1}{dt}\right)^2dt+bx_1^2(t)=0.$$

这个等式只能当 $x_1(t)\equiv 0$ 时成立。否则我们将有方程的左边为正数，而右边为零，当 $t>t_0$ 时，这是个矛盾。同样可以研究 $t<t_0$ 的情形。

为了要对任意的常系数 m, a, b, 来证明我们的断言，我们来看函数 $y_1(t)=x_1(t)e^{-\alpha t}$，不难验算这个函数满足初始值为零的条件。而当 $\alpha>0$ 取 α 相当大则可以使得函数 $y_1(t)$ 满足形如(6)的方程并且 $a>0$, $b>0$, $m>0$。将 $x_1(t)=y_1(t)e^{\alpha t}$ 及其微商代入(6)便容易得到这个方程。因此，由前面已证实了的，有 $y_1(t)\equiv 0$，即 $x_1(t)=y_1(t)e^{\alpha t}\equiv 0$。

因此，我们证明了公式(20)，(21)，(23)给出了方程(6)的所有的解。

我们来看看，由这些公式可以得到方程(6)的解的那些性质。

用

$$\lambda_{1,2}=-\frac{a}{2m}\pm\sqrt{\frac{a^2}{4m^2}-\frac{b}{m}} \tag{25}$$

表方程(19)的根．由导致方程(6)的物理例子可设 $m>0$, $a\geqslant 0$, $b>0$.

情形 1 $a^2>4bm$．示性方程(19)的两根为实的、负的且不相等．在此情形，公式(20)所给出的函数 $x(t)$ 表示方程(6)的通解．这个公式所表示的函数和它的微商，当 $t\to\infty$ 时，趋于零，并且不多于一个 t 值使这个函数为零，也不多于一个 t 值使它的微商为零．因此函数 $x(t)$ 不多于一个极大值或极小值．物理意义便是，由于介质的阻尼如此之大，致使振动不会产生．运动的质点穿过平衡位置 $x=0$ 不能多于一次．此外，质点达到离 $x=0$ 的点某个最远的位置，然后开始慢慢地向平衡位置返回，但不再穿过它．

情形 2 $a^2=4bm$．方程(19)的两根相等，公式(21)表示了方程(6)的通解．在这种情形下，当 $t\to+\infty$ 时，函数 $x(t)$ 及其微商 $x'(t)$ 也趋于零．此时 $x(t)$ 及 $x'(t)$ 不能多于一次为零．具有横坐标 $x(t)$ 的运动质点的性质和情形 1 相同．

情形 3 $a^2<4bm$．示性方程(19)的根的虚数部分不为零．公式(23)给出方程(6)的通解．对于方程(6)的所有的解而言，质点 x 在 x 轴方向以同一不变的周期 $\frac{2\pi}{\beta}$ 振动，其振幅为 $Ce^{\alpha t}$，其中 $\alpha=-\frac{a}{2m}$.

不具外力干扰下，物理系统的振动称为这个系统的**固有振动**．由前面的例子 2, 3, 4, 5 可见这种系统的固有振动的周期只与系统的结构有关．对于这些系统可能产生的所有的固有振动周期都是相同的．在例 2 中周期等于 $2\pi:\sqrt{\frac{b}{m}-\frac{a^2}{4m^2}}$；在例 4 中周期等于 $2\pi:\sqrt{\frac{kps}{v\rho l}}$；在例 5 中周期等于 $2\pi:\sqrt{\frac{1}{LC}-\frac{R^2}{4L^2}}$.

如果 $a=0$，也即是介质对运动没有阻力，则振动的振幅是常数．质点作谐振运动．而当 $a>0$，即介质对运动有阻力，即使

阻力小于($a^2<4bm$)，振动的振幅趋于零，而运动衰减。

最后，在任何情况下，方程(6)的解 $x(t)\equiv 0$ 都表示在任何时间处于位置 $x=0$ 的质点 x 的平衡状态，这点我们称为平衡点.

如果方程(19)的两根的实数部分为负的，则由公式(20)，(21)，(23)所给出的解. 可见，当 $t\to+\infty$，则这些解及其微商都趋于零，也即是，随着时间的增加，振动衰减下去.

而当方程(19)即使有一个根具有正实数部分的情形时，则方程(6)将有当 $t\to+\infty$ 时不趋于零的解，则当 $t\to+\infty$ 时方程(6)的某些解将为无界，如果 $m>0$，则这种情形只有当 b 为负数或 a 为负数的时候才有可能. 物理意义上，这种情形对应于弹性力不是将质点 x 吸返平衡位置而是推离平衡位置的情形或者介质的阻力是负的情形. 这些情形虽然在本章开始时所举的例子中不存在，但它们在其他的物理模型中是存在的.

如果方程(19)的根 λ_1 与 λ_2 的实数部分为零，这种情形只有当方程(19)中 a 等于零时才有可能，则由公式(23)，当 $\alpha=0$ 时，质点 $x(t)$ 作有限振幅及有限速度的谐振运动.

常系数非齐次线性方程 我们详细地来研究方程

$$m\frac{d^2x}{dt^2}+a\frac{dx}{dt}+dx=A\cos\omega t, \tag{26}$$

这是在弹性力、介质阻力及周期外力 $A\cos\omega t$ 作用下，质点的线性振动方程[参考 §1 方程(6′)].

方程(26)称为非齐次线性方程. 方程(6)是它对应的齐次方程.

现在来求方程(26)的通解.

注意到，非齐次方程的解与对应于它的齐次方程的解的和仍然是非齐次线性方程的解. 因此，要找方程(26)的通解只要知道这个方程的任何一个特解便够了. 方程(26)的通解可表为其一特解与其对应的齐次方程的通解的和.

自然可以希望，对于周期外力，运动也有同样的周期才合拍，故来找方程(26)具有形式 $x=B\cos(\omega t+\delta)$ 的特解，其中 B 与 δ 是某些常数. 要决定 B 与 δ 使得函数

$$x=B\cos(\omega t+\delta)$$

满足方程(26)计算其微分 $\frac{dx}{dt}=-B\omega\sin(\omega t+\delta)$，及 $\frac{d^2x}{dt^2}=-B\omega^2\cos(\omega t+\delta)$，并将结果代入(26)，得 $m[-B\omega^2\cos(\omega t+\delta)]+a[-B\omega\sin(\omega t+\delta)]+bB\cos(\omega t+\delta)=A\cos\omega t$. 由大家熟知的公式得到

$$B[(b-m\omega^2)\cos(\omega t+\delta)-a\omega\sin(\omega t+\delta)]$$
$$=B\sqrt{(b-m\omega^2)^2+a^2\omega^2}\cos(\omega t+\delta')=A\cos\omega t,$$

其中 $\delta'=\delta+\gamma$，及 $\gamma=\operatorname{arc\,tg}\frac{a\omega}{b-m\omega^2}$. 显然，如果我们命

$$\delta=-\operatorname{arc\,tg}\frac{a\omega}{b-m\omega^2}\text{ 及 }B=\frac{A}{\sqrt{(b-m\omega^2)^2+a^2\omega^2}},$$

则函数 $x=B\cos(\omega t+\delta)$ 将满足方程(26).

当 $(b-m\omega^2)^2+a^2\omega^2\neq0$，则 $B\cos(\omega t+\delta)$ 形式的解必定存在. 在 $(b-m\omega^2)^2+a^2\omega^2=0$ 的情形，即 $a=0$ 与 $b=m\omega^2$ 时，则方程(26)具有以下的形式

$$m\frac{d^2x}{dt^2}+m\omega^2x=A\cos\omega t.$$

很容易验证，此时，函数 $x=\frac{At}{2\sqrt{mb}}\sin\omega t$ 为特解。

非齐次方程(26)的解特称为强迫振动，因子

$$\varphi(\omega)=\frac{1}{\sqrt{(b-m\omega^2)^2+a^2\omega^2}}$$

表示我们所要求的强迫振动的振幅 B 与扰动力的振幅 A 的比. 函数 $\varphi(\omega)$ 所表示的曲线称为共振曲线，使 $\varphi(\omega)$ 达到最大值时的频率 ω 称为共振频率. 现在来求它. 如果 $\varphi(\omega)$ 在 $\omega_1\neq0$. 取最大值，则在 ω_1，$\varphi'(\omega)$ 变成零，即是

$$-4(b-m\omega_1^2)m\omega_1+2a^2\omega_1=0,$$

因此

$$\omega_1=\sqrt{\frac{b}{m}-\frac{a^2}{2m^2}}.$$

此时

$$\varphi(\omega_1)=\frac{1}{a\sqrt{\frac{b}{m}-\frac{a^2}{4m^2}}}.$$

由此可见，当 a 越小，强迫振动在 $\omega=\omega_1$ 的振幅越大. 对很小的 a，频率 ω_1 接近于数值 $\sqrt{\frac{b}{m}}$，即是自由振动的频率. 当 $a=0$ 及 $b=m\omega^2$，则我们已看到，强迫振动有以下形式：

$$x=\frac{At}{2\sqrt{mb}}\sin\omega t,$$

即当 $t\to\infty$，这个振动的振幅无限增加. 这种现象在数学上称为共振. 当外力的周期与系统的固有振动的周期一致时，共振便产生了. 在实际情形中，当外力的周期与固有振动的周期很接近时，系统的振动范围可能非常大.

在系统中产生巨大振动的可能性常常用于制造无线电技术中的各种放大器方面. 但是大振动也可能导致某些建筑物，例如桥梁及横跨的建筑的破坏. 因此，预见到共振或接近于共振的振动的发生的可能性是很重要的.

前面已提过，方程(26)的任何解可以由我们所求得的强迫振动和用公式(20)，(21)，(23)所表示的齐次方程的一个解的和表示出来. 当 $a>0$ 及 $b>0$ 时，齐次方程的解在 $t\to\infty$ 时趋于零；即随着时间的增加，何任运动将趋近于我们所得到的强迫振动. 如果 $a=0$，$b>0$，则强迫振动与系统的不衰减的固有振动叠加起来. 当 $b=m\omega^2$ 及 $a=0$ 共振发生了.

如果在系统上施加某些周期外力 $f(t)$，则可以用以下方法来求得系统的强迫振动. 在足够的准确度上，将 $f(t)$ 表为三角级数的一段[1)]

$$\sum_{i=1}^{n}(a_i\cos\omega_i t+b_i\sin\omega_i t).\tag{27}$$

对这个和数的每一项求强迫振动. 将(27)的每项所得的强迫振动加起来，便得到对应于力 $f(t)$ 的强迫振动. 如果频率 ω_k 中有些与系统的固有振动的频率重合，则共振便发生了.

1) 参看第十二章，§7.

§3. 微分方程的解及应注意的几个方面

象常系数线性方程那样的所有的解可以用很简单的函数明显地表示出来的微分方程是不多的. 可以举出很简单的微分方程的例子，其通解不能由已知函数用有限个数的积分表示出来，或者说不能表示为积分.

在 1841 年，刘微尔已经证明，李嘉蒂型方程 $\frac{dy}{dx}+ay^2=x^2$ 的解不能用初等函数的有限个积分来表示，其中 $a>0$. 因此，能应用于广泛类型的微分方程的近似解的方法具有很大的意义.

找不到方程的准确解而只得到近似解这个事实，不应使我们感到不安. 首先，至少在原则上这种近似解可能达到任何高度的准确度. 其次，值得特别注意的是，在大部分情形，用来描述这个或那个物理过程的微分方程本身也不是完全准确的. 从 §1 中所举出的各个例子便可以看出这点.

发声共振器方程(12)在这方面表现得最明显. 在导出这个方程的时候，我们略去了容器颈部空气的压缩性以及容器中空气的运动. 其实当颈部的空气在运动时，容器中的空气也在运动，但是这两种运动的速度与位移是不相同的. 空气质点的位移在颈部比在容器中大. 因为我们略去了在容器中的空气移动，而只考虑其压缩性. 对于颈部的空气则相反，略去了压缩的能量而只考虑它的运动的动能.

当导出物理摆的微分方程时我们略去了链子的质量，即系着摆的链子看作没有质量的. 当导出回路中的电振荡方程(14)时我们略去了导体的自感，线圈的电阻. 一般说来，当导出任何物理系统的微分方程时，我们略去一些因素而理想化了它. 因此安德洛诺夫对这件事引起很大的注意，即只有当方程经过任何意义下的小变动时，其解也变动得很小，这种微分方程对于物理的研究才有特别的兴趣. 他称这种方程是"粗糙的". 这种方程特别值得全面

的研究。

应当指出，在研究物理过程中，用微分方程来描述物理量的变化的规律，不只这个方程本身不完全准确，甚至于这些量的个数也只是很近似地定出的。例如，严格地说就没有绝对的刚体。因此当研究摆的振动时，我们应该考虑到系着摆的链子的变形，也应该考虑到，当作是质点的刚体本身的变形。同样在研究系于弹簧上的重物的振动时，必须要考虑到弹簧的每一圈的质量。但在这些例子中容易证明，摆或系重物的弹簧的每个个别质点的运动性质对于振动的性质很少影响。如果考虑这些影响，问题便很复杂，而不能得到很好的近似解。对应于物理的现实而言，这样得到的解并不优于前面§1中不考虑这些影响所得的解。问题的这些理想化总是不可避免的。在描述物理过程时，要考虑到过程的基本特点，而不能无例外地考虑到所有的各方面。这不仅会使问题复杂化，而且在大部分情形使得问题的解决成为不可能。在研究某一个过程时，物理及力学的问题在于找出尽可能少的量的个数，只要足够确切地决定在每个时间所研究的过程的状态；找出尽可能简单的微分方程，只要能很好地描述这些量的变化的规律。这种问题常常是很不容易的。在研究某一物理问题时，哪些不可省略，哪些可以省略是很关本质的，归根结底也只有用长时期的实验来解决。只有当我们经过理想化的研究所得的答案和实验的结果符合时，才能认为这个规律是被理想化的。

关于量的个数减少的可能性的数学问题，最简单而又示性的提法之一可叙述如下：

设我们首先对所研究的物理系统的状态在 t 时用两个变量 $x_1(t)$ 与 $x_2(t)$ 来表征它。设决定它的变化的规律的微分方程有以下的形式

$$\frac{dx_1}{dt}=f_1(t,\ x_1,\ x_2),$$

$$\varepsilon\frac{dx_2}{dt}=f_2(t,\ x_1,\ x_2). \tag{28}$$

在第二个方程中，微商的系数是一个小的常数参数 ε. 如果我们置 $\varepsilon=0$，则(28)的第二方程不再是微分方程. 它变成以下形式

$$f_2(t,\ x_1,\ x_2)=0.$$

由此决定 x_2 为 t 及 x_1 的函数，代入方程(28)中的第一式，我们便得到只有一个变量 x_1 的微分方程

$$\frac{dx_1}{dt}=F(t,\ x_1).$$

因此，在这种研究下变量的数目降到一个. 现在要问，在什么条件下，我们取 $\varepsilon=0$，所引起的误差很小. 可能出现这种情形，当 $\varepsilon\to0$ 时，$\frac{dx_2}{dt}$ 的量无限增大，使方程(28)的第二方程的右边，当 $\varepsilon\to0$ 时，不趋近于零. 类似于上面所提出的一系列的问题，苏联数学家们已给了很完整的解答.

§4. 微分方程积分问题的几何解释. 问题的推广

为简单计，开始我们只研究一个具有一个未知函数的一阶微分方程

$$\frac{dy}{dx}=f(x,\ y),\tag{29}$$

式中函数 $f(x,\ y)$ 在 $(x,\ y)$ 平面上的某一区域 G 中定义的. 过这个区域中每一点，这方程的解的图形的切线的角系数即由方程(29)所规定. 假定在区域 G 中的每一点上，用一线段来表示在这一点由 $f(x,\ y)$ 的值所决定的切线的方向(这线段上正向或负向对我们均可采取)，则得到一个方向场. 则在初值条件 $y(x_0)=y_0$ 下微分方程(29)求解的问题可以归结如下：在区域 G 中要找穿过点 $M_0(x_0,\ y_0)$ 的曲线 $y=\varphi(x)$，这曲线上每一点的切线的方向是(29)在这点所定的值，或者简单地说，这曲线上的每一点有预先给定的方向.

从几何的观点来看，在下面的一些情形中，上述的问题的提法

有些不自然的地方:

1. 为使得在区域 G 中任何一点 (x, y) 其角系数是 $f(x, y)$，我们必须除掉平行于 Oy 的方向，因为一般说来我们只研究有限的值；特别，我们需要假定方程(29)的右边的函数 $f(x, y)$ 在所有的点上只有有限值.

2. 所研究的曲线只是 x 的函数的图形，因此，如果某一曲线它与垂直于 Ox 轴的另一直线有多于一点的交点，这种曲线我们便必须除掉，因为我们研究的曲线是由 $y=\varphi(x)$ 而定的，这对一个 x 只有一个 y 值，即为 x 的单值函数；因此，又需要假定这个微分方程的所有的解都是 x 的单值函数.

由上面这些缺点看来，我们需要把微分方程(29)的求解问题的上述提法加以推广，即，需要允许在某些点的方向场可以平行于 Oy 轴. 在这些点，对于 Ox 轴的角系数没有意义，而要利用对于 Oy 轴的角系数. 对应于这种情形，除了考虑方程(29)还要考虑方程

$$\frac{dx}{dy}=f_1(x, y), \tag{29'}$$

式中 $f_1(x, y)=\dfrac{1}{f(x, y)}$，当 $f(x, y)\neq 0$. 当方程(29)没有意义而方程(29′)有意义时，便要用方程(29′). 微分方程(29)与(29′)的积分问题可以提出如下：在区域 G 中找出所有的线，这些线上每一点的方向场由方程(29)及(29′)所给出. 这些曲线我们称为方程(29)及(29′)或由方程(29)及(29′)所决定的方向场的积分线(积分曲线). 以后经常用单数的"方程(29)，(29′)"来代替多数的"方程(29)，(29′)"[1]. 显然，方程(29)所有的图形都是方程(29)，(29′)的积分曲线. 但并非所有的方程(29)，(29′)的积分曲线都是方程(29)的解的图形. 例如当这种线与垂直于 Ox 轴的某一直线有多于一个的交点时，便是如此.

以下，如果已知

---

* 俄文文法上的简化，中文没有分别. ——译者注

$$f(x,\ y)=\frac{M(x,\ y)}{N(x,\ y)},$$

则对于方程
$$\frac{dy}{dx}=\frac{M(x,\ y)}{N(x,\ y)},$$

我们便不用方程

$$\frac{dx}{dy}=\frac{N(x,\ y)}{M(x,\ y)}.$$

有时引入参数 t, 我们把这些方程写成方程组

$$\frac{dx}{dt}=N(x,\ y),\quad \frac{dy}{dt}=M(x,\ y),$$

其中 x 与 y 看作 t 之函数.

例 1 方程

$$\frac{dy}{dx}=\frac{y}{x}, \tag{30}$$

除了原点外, 给出了方向场. 这个方向场的图形如图 7. 方程(30)所决定的所有的方向都经过原点. 显然, 对任何 k, 函数

$$y=kx \tag{31}$$

都是方程(30)的解. 这个方程的积分线的全体可以由关系

$$ax+by=0 \tag{32}$$

表出, 其中 a, b 为不同时为零的任意常数. Oy 轴是方程 (30) 的积分线, 但不是它的解的图形.

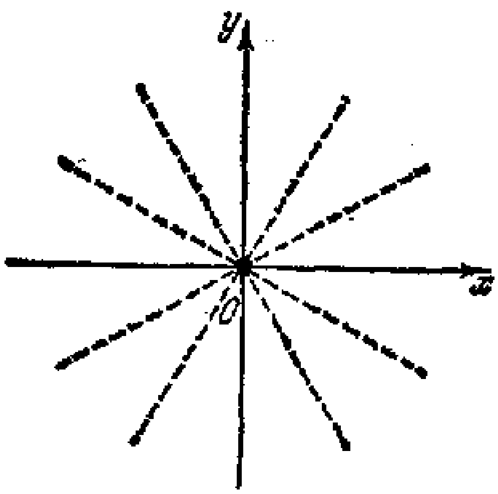


图 7

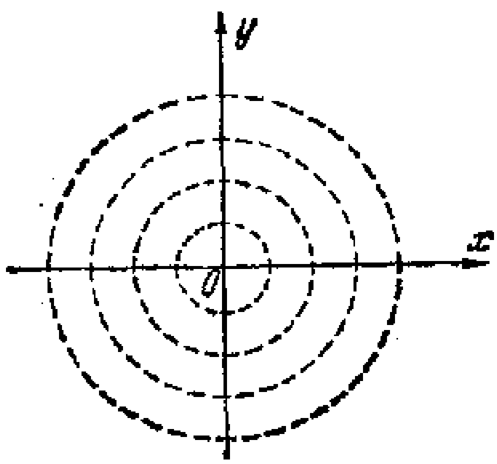


图 8

在坐标原点处方程(30)不定出方向场，因此，严格地说，线与(32)除掉原点之外才处处是积分线．因此，正确地说，方程的积分线不是通过坐标原点之直线，而是由坐标原点射出的直线．

例 2 方程

$$\frac{dy}{dx}=-\frac{x}{y} \tag{33}$$

除了坐标原点外处处都定义了方向场．它的图形如图 8．在任何一点(x, y)，方程(30)与(33)所决定的方向互相垂直．显然所有圆心在坐标原点的圆都是方程(33)的积分曲线．这个方程的解又可写成两函数

$$y=+\sqrt{R^2-x^2},\ y=-\sqrt{R^2-x^2},\ -R\leqslant x\leqslant R.$$

以后，为简便计，对于严格地说来是"经过点(x, y)的解的图形"我们只称为"经过点(x, y)的解"．

§5．微分方程解的存在性与唯一性方程的近似解

微分方程解的存在性与唯一性的问题 我们回到 n 阶微分方程(17)．一般说来它有无限多个解，要从它的所有可能的解中，找出某一个确定的解，必须给方程加上补充条件，条件的个数应等于方程的阶数 n．这类条件可以有各种各样的性质，其形式则与导出这个方程的物理、力学或其他的问题密切相关．例如，如果我们要研究力学系统的运动，设其开始时具有确定的初始状态，则附加的条件将是与独立变量的确定的(初始)值有关，称为问题的初始条件．如果我们想定出悬桥建筑的悬挂下的曲线或者我们要研究两端支持住不负重物的梁的弯曲，我们遇到的附加条件与独立变量的不同的值有关(在我们的例中悬桥的两端点及梁的两支点)．可以举出很多例子说明附加于微分方程的条件可以非常多样性的．

设给定附加条件，我们要找满足这些条件的方程(17)的解。我们要提出的第一个问题是所求的解的存在性问题．我们常是不能断定它的存在性．例如设方程(17)描述了某个物理试验的工作规律，我们想确定，在这个试验中是否存在周期过程．附加的条件便是，在这个试验中周期地回到过程的初始状态，我们预先不能说是否存在满足这个条件的解．

在所有的情形中，解的存在性与唯一性问题的研究使得我们有可能说明，附加那些条件对我们所研究的方程是可以满足的，那些唯一地确定了解．对于描述某个物理现象的微分方程，指明解的存在性与唯一性的条件，并加以证明，对于这个物理理论本身，具有很大的意义，它指明了，描述这个现象的数学所采取的假设彼此不相矛盾和这个描述的完整性．

存在性问题的研究方法是多种多样的，但其中特别重要的是所谓直接法．这类方法中要求的解的存在性是用作近似解的方法来证明的，这些近似解的极限收敛于我们的问题所求的确切解．这类方法不只是能够建立确切解的存在性，而且还有可能，至少从理论上，任意准确地接近它．

在本节下面为了明确起见，我们只研究具有已知初始值的问题，对这个问题设法说明欧拉方法和逐次逼近法的思想．

欧拉折线法　设在平面(x, y)上的某区域G中给定微分方程

$$\frac{dy}{dx}=f(x, y). \tag{34}$$

我们已经指出过，方程(34)在区域G中定义一方向场．在G中任取一点(x_0, y_0)．过这一点作一角系数为$f(x_0, y_0)$的直线L_0，这直线切于过(x_0, y_0)点的积分线．在直线L_0上任取一点(x_1, y_1)，相当接近于(x_0, y_0)［在图9上(x_1, y_1)点用1记］．过(x_1, y_1)作直线L_1，其角系数为$f(x_1, y_1)$，在其上取一点(x_2, y_2)(在图9上这点用2记)过(x_2, y_2)取直线L_2，在其上取(x_3, y_3)等等．设$x_0<x_1<x_2<x_3<\cdots$，自然也设所有这些点(x_0, y_0)，(x_1, y_1)，(x_2, y_2)，…都在G内．连接这些点的折线叫做欧拉折线．也可以在

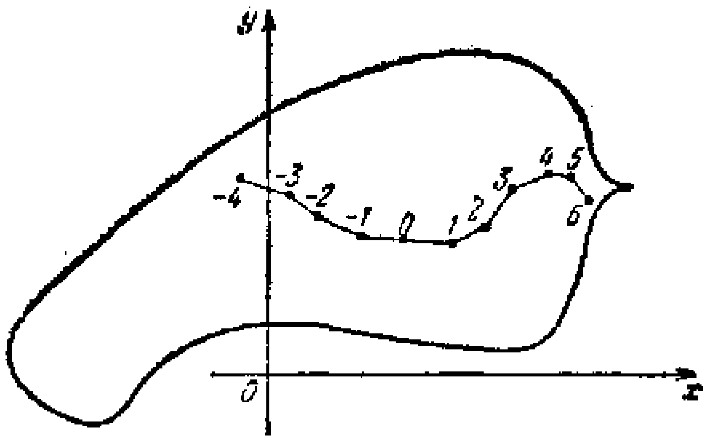


图 9

x 减少的方面作欧拉折线（在图 9 上对应的角点用 -1, -2, -3 表之）.

自然希望，通过(x_0, y_0)点的每一条欧拉折线，当每段都很短时，可以作为通过(x_0, y_0)点的积分曲线 l 的某种表示，当最长的段都趋近于零时，即每段也趋近于零时，欧拉折线接近于这条积分曲线.

在此时我们自然假定这个积分线存在. 事实上，不难证明，如果函数 $f(x, y)$ 在区域 G 内连续，则可以选出无限序列的欧拉折线，其最长的直线段趋近于零，而这个序列收敛于某个积分曲线 l. 但是在此时，一般说还不能认为是唯一的：可能存在不同序列的欧拉折线，它们收敛于不同的积分曲线，均穿过同一点(x_0, y_0). 拉夫伦捷夫作出了这样的例子，具有连续函数 $f(x, y)$ 的 (29) 形式的微分方程，在区域 G 中任何一点 P 的任何邻域中，不只有一条而至少有两条积分线穿过 P 点. 为要使得经过区域 G 中的每一点只有一条积分线，对函数 $f(x, y)$ 必须具有比连续性更多的补充条件. 例如，假定函数 $f(x, y)$ 连续，而且在 G 中对 y 有有界微商便足够了. 在这种情形下，可以证明，经过 G 内的每一点存在一条而且只有一条积分线，而且当折线的最长线段趋近于零时，过(x_0, y_0)点的所有的欧拉折线序列一致收敛于这一条唯一的积分线. 因此，当折线的线段足够小时，欧拉折线便可近似地取作方程 (34) 的积分曲线.

由上可见，欧拉折线的作法是用切于积分线的直线段来代替积分线的小段的弧．在实际中常不是用切于积分线的直线段去作成对微分方程(34)的积分线的近似，而是用与积分曲线有更高阶相切的抛物线的线段来作近似．由此可以用更少的步骤来得到有同样准确度的近似解(即作近似线时用更少的曲线段)．抛物线

$$y=a_0+a_1(x-x_k)+a_2(x-x_k)^2+\cdots+a_n(x-x_k)^n, \qquad (35)$$

在(x_k, y_k)点与方程(34)过此点的积分线有 n 阶相切，则其系数由公式

$$a_0=y_k, \qquad (36)$$

$$a_1=\left(\frac{dy}{dx}\right)_{x=x_k}=f(x_k, y_k), \qquad (36')$$

$$\begin{aligned}2a_2&=\left(\frac{d^2y}{dx^2}\right)_{x=x_k}=\left[\frac{df(x, y)}{dx}\right]_{x=x_k}\\&=f'_x(x_k, y_k)+f'_y(x_k, y_k)\left(\frac{dy}{dx}\right)_{x=x_k}\\&=f'_x(x_k, y_k)+f'_y(x_k, y_k)f(x_k, y_k),\end{aligned} \qquad (36'')$$

$$\begin{aligned}6a_3&=\left(\frac{d^3y}{dx^3}\right)_{x=x_k}\\&=\left\{\frac{d}{dx}\left[f'_x(x, y(x))+f'_y(x, y(x))f(x, y(x))\right]\right\}_{x=x_k}\\&=f''_{xx}(x_k, y_k)+2f''_{xy}(x_k, y_k)f(x_k, y_k)\\&\quad+f''_{yy}(x_k, y_k)f^2(x_k, y_k)+f'^2_y(x_k, y_k)f(x_k, y_k)\\&\quad+f'_y(x_k, y_k)f'_x(x_k, y_k)\end{aligned} \qquad (36''')$$

等等来确定．我们只要用多项式(35)来计算在 $x=x_{k+1}$ 的值．我们不需要知道系数 $a_0, a_1, a_2, \cdots, a_n$ 本身的值．存在很多方法来计算当 $x=x_{k+1}$ 时多项式(35)的值，但是可以避免计算由(36)所决定的系数 $a_0, a_1, \cdots, a_n$.

还有基于其他思想的找寻微分方程(34)的解的近似方法．可行方法中的一个是克雷洛夫(1863—1945)院士所发展的．

逐次逼近法 现在我们来谈谈与欧拉折线法有同样广阔应用范围的逐次逼近法．仍然假定我们所要找的微分方程(34)的解

$y(x)$满足初始条件

$$y(x_0)=y_0.$$

任取一个函数 $y_0(x)$作为解 $y(x)$的开始的近似. 为了简便起见，假定它满足初始条件，只是这假定并不是必要的. 代入方程右边 $f(x, y)$中的未知函数 y,用以下的方法作解 y 的第一次近似

$$\frac{dy_1}{dx}=f(x, y_0(x)), \quad y_1(x_0)=y_0.$$

因为方程右方的函数为已知,所以 $y_1(x)$可用积分求出

$$y_1(x)=y_0+\int_{x_0}^{x} f(t, y_0(t))dt.$$

可以希望 $y_1(x)$与解 $y(x)$的差比 $y_0(x)$与解 $y(x)$的差小，这是因为在作 $y_1(x)$时，我们用了微分方程,可能它将初始近似的误差改正了一些. 也可以看到，如果用同样方法把第一次近似 $y_1(x)$ 改善一下,则第二次近似

$$y_2(x)=y_0+\int_{x_0}^{x} f(t, y_1(t))dt$$

将更接近于所求的解.

假设,这种改善的过程可以无限继续下去,则作出一系列的近似

$$y_0(x),\ y_1(x),\ \cdots,\ y_n(x),\ \cdots.$$

它们是否收敛于解 $y(x)$呢?

更详细的研究证明,当函数 $f(x, y)$连续及 f'_y 在区域 G 中为有界,则至少对于离 x_0 不远之 x,函数 $y_n(x)$真正地趋近于确切解 $y(x)$. 如果我们反复计算足够多次，则我们可以得到具有任何准确度的解 $y(x)$.

象我们寻找方程(34)的近似积分线一样，我们可以找两个或多个一阶微分方程的方程组的积分线. 只要我们能从这些方程把未知函数的微商解出来. 设,例如已知方程组

$$\frac{dy}{dx}=f_1(x, y, z), \quad \frac{dz}{dx}=f_2(x, y, z). \tag{37}$$

设这些方程的右边为连续的，又对 y 及对 z 在某一空间区域

G 内有有界偏微商．在这些条件下，可以证明，在方程(37)的右边部分的定义区域 G 中，经过每一点(x_0, y_0, z_0)有一条而且只有一条方程组的积分线

$$y=\varphi(x),\ z=\psi(x).$$

函数 $f_1(x, y, z)$与 $f_2(x, y, z)$是过(x, y, z)点的积分线在这点的切线的角系数．为了近似地求得 $\varphi(x)$与 $\psi(x)$可以用欧拉折线法或类似于应用到方程(34)的其他方法．

由初始条件计算常微分方程近似解的过程可以用在计算机上．有工作得很快的计算机，例如，如果计算程序已经作好，机器已准备计算，计算机算出炮弹弹道的时间要比炮弹飞到弹着点的时间短得多(参看第十四章)．

高阶微分方程与多个一阶微分方程组的联系 如果对于每个未知函数的最高阶微商可以解出来的常微分方程组，一般说来都可以用引入新的未知函数把它们化为对所有的微商都解出来的一阶方程组．例如，设已知微分方程

$$\frac{d^2y}{dx^2}=f\left(x,\ y,\ \frac{dy}{dx}\right). \tag{38}$$

置

$$\frac{dy}{dx}=z. \tag{39}$$

则方程(38)可以写成以下形式

$$\frac{dz}{dx}=f(x,\ y,\ z). \tag{40}$$

这样方程(38)的每个解对应于由方程(39)及(40)所组成的方程组的某个解．也容易看出，方程组(39)及(40)的每个解也对应于方程(38)的解．

不包含独立变数的方程 在§1中所看到的摆的问题，赫尔姆霍尔茨的声学共振器的问题，最简单的电回路的问题，真空管振荡器的问题，都化为独立变数(时间 t)不明显出现的微分方程．现在我们来研究这种类型的二阶方程的问题，但对于这种类型的方程我们可以把它化为一个一阶的微分方程来研究，并不是象前面

所说的那样，一个一般的二阶方程可以化成一个方程组. 这样使得研究大为方便.

因此，设给定不包含时间 t 的二阶微分方程

$$F\left(x,\ \frac{dx}{dt},\ \frac{d^2x}{dt^2}\right)=0. \tag{41}$$

记

$$\frac{dx}{dt}=y, \tag{42}$$

又把 y 看成 x 之函数，那么

$$\frac{d^2x}{dt^2}=\frac{d}{dt}\left(\frac{dx}{dt}\right)=\frac{dy}{dt}=\frac{dy}{dx}\cdot\frac{dx}{dt}=y\frac{dy}{dx}.$$

因此方程(41)可以写成形式

$$F\left(x,\ y,\ y\frac{dy}{dx}\right)=0. \tag{43}$$

这样，方程(41)的每个解对应于方程(43)的一个解. 方程(43)的每个解 $y=\varphi(x)$ 则对应于方程(41)的无穷多个解. 这些解可以由方程

$$\frac{dx}{dt}=\varphi(x) \tag{44}$$

积分得到，其中 x 看作 t 的函数.

显然，如果 $x=x(t)$ 为满足此方程的任一解，则对任何常数 t_0，函数 $x(t+t_0)$ 也是满足这个方程的.

也可能出现这样的情形，并非方程(43)的所有的积分线都是一个 x 的函数的图形. 例如有闭曲线时便是如此. 那么，方程(43)的积分线应该分为若干段弧，其中每一段都是 x 的函数的图形. 对每一段，都要解方程(44).

对应于方程(41)的物理系统的状态，可以用量 x 及 $\frac{dx}{dt}$ 来表示任何时候的特征，这些量称为系统的相. 对应于这些的平面(x, y)称为方程(41)的相平面. 这个方程的每个解 $x=x(t)$ 对应于平面(x, y)上的曲线

$$x=x(t),\ y=x'(t);$$

这里把 t 看作参数．反之：在平面 (x, y) 上方程(43)的每一条积分线 $y=\varphi(x)$ 对应于方程(41)的形如 $x=x(t+t_0)$ 的无限多个解；式中 t_0 是任意常数．在平面上有了方程(43)的积分线的分布的全部图形时，不难表示出方程(41)的所有可能的解的性质，例如方程(43)的每一条闭积分线对应于方程(41)的周期解．

用变换(42)将方程(6)化为方程

$$\frac{dy}{dx}=\frac{-ay-bx}{my}. \tag{45}$$

以 $v=x$ 及 $\frac{dv}{dt}=y$ 代入方程(16)我们同样得到

$$L\frac{dy}{dx}=\frac{-[R-M(a_1+2a_2x+3a_3x^2]y-x}{y}. \tag{46}$$

二阶方程(41)所对应的物理系统的每个时间的状态可以用两个量(相)[1] x 与 $\frac{dx}{dt}$ 来表示，同样的，对于高阶方程或方程组所描述的物理系统的状态可以由几个量(相)来表示．前面所讲的是相平面，这时就讲相空间了．

§6. 奇 点

在区域 G 中研究方程

$$\frac{dy}{dx}=\frac{M(x, y)}{N(x, y)}. \tag{47}$$

设 $P(x, y)$ 点在 G 内．

如果可能指出 P 点的这样的邻域 R，经过 R 中的每点有一条而且只有一条方程(47)的积分线，则我们称 P 点为方程(47)的常点．如果找不出这样的 R，则我们称 P 点为这方程的奇点．奇点的研究对于微分方程定性理论有很大的意义，下节我们将讨论定性理论．

1) 这时 $\frac{d^2x}{dt^2}$，$\frac{d^3x}{dt^3}$ 由方程(41)及微分公式(36)可由 x 及 $\frac{dx}{dt}$ 的值算出．

一个奇点如果它有一个邻域，在其中没有其他的奇点则这种奇点称为**孤立奇点**. 孤立奇点有特别重大的意义. 当研究(47)形式的方程时，设 $M(x, y)$ 与 $N(x, y)$ 对 x 与 y 有高阶偏微商，这种情形在应用中经常遇到. 对于这样的方程，区域中任一个内点，只要 $M(x, y)\neq 0$ 或 $N(x, y)\neq 0$，都是常点. 考虑任一个这样的内点 (x_0, y_0)，在这点上有 $M(x, y)=N(x, y)=0$，为简便起见，把这点记作 $x_0=0$ 与 $y_0=0$. 只要把坐标原点移到 (x_0, y_0) 上去，便可以达到简化的目的. 将 $M(x, y)$ 与 $N(x, y)$ 用台勒级数依 x 与 y 之幂展开，并只考虑一次的项，则在 $(0, 0)$ 点的邻域内，我们得到

$$\frac{dy}{dx}=\frac{M'_x(0, 0)x+M'_y(0, 0)y+\varphi_1(x, y)}{N'_x(0, 0)x+N'_y(0, 0)y+\varphi_2(x, y)}, \tag{48}$$

式中 $\varphi_1(x, y)$ 与 $\varphi_2(x, y)$ 是 x 与 y 的某种函数，使得

$$\lim_{\substack{x\to 0\\ y\to 0}}\frac{\varphi_1(x, y)}{\sqrt{x^2+y^2}}=0 \quad 及 \quad \lim_{\substack{x\to 0\\ y\to 0}}\frac{\varphi_2(x, y)}{\sqrt{x^2+y^2}}=0.$$

方程(45)与(46)便是这种类型. 方程(45)在 $x=0$ 与 $y=0$ 既不确定 $\frac{dy}{dx}$ 也不确定 $\frac{dx}{dy}$. 如果行列式

$$\begin{vmatrix} M'_x(0, 0) & M'_y(0, 0) \\ N'_x(0, 0) & N'_y(0, 0) \end{vmatrix}\neq 0,$$

则不论我们在原点给 $\frac{dy}{dx}$ 任何值，对于 $\frac{dy}{dx}$ 及 $\frac{dx}{dy}$ 的值在原点都是间断点. 因为，当沿不同的路线接近原点时，可以得到不同的极限值. 对于我们的微分方程，坐标原点是奇点.

已经证明，只要方程

$$\begin{vmatrix} \lambda-M'_y(0, 0) & -M'_x(0, 0) \\ -N'_y(0, 0) & \lambda-N'_x(0, 0) \end{vmatrix}=0 \tag{49}$$

的两根的实数部分都不等于零，则在方程(48)中分母与分子中的函数 $\varphi_1(x, y)$ 与 $\varphi_2(x, y)$ 对于孤立奇点(在我们的情形是坐标原点)附近的积分曲线的分布的性质没有什么影响. 因此，要得到这种分布的概貌，我们只要研究方程

$$\frac{dy}{dx}=\frac{ax+by}{cx+dy} \tag{50}$$

在坐标原点附近积分线的分布情形，其中行列式

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0.$$

我们注意到，积分曲线在微分方程奇点附近的分布性质，对于很多力学问题，例如在平衡位置附近运动的轨线的研究，具有很大的价值。

已经证明，在平面上总可以找到与 x 及 y 用等式

$$\begin{aligned} x&=k_{11}\xi+k_{12}\eta, \\ y&=k_{21}\xi+k_{22}\eta \end{aligned} \tag{51}$$

所联系的坐标 ξ 及 η，其中 k_{ij} 是实数，使得方程(50)化为下面三种类型之一：

1) $$\frac{d\eta}{d\xi}=k\frac{\eta}{\xi}，\quad 其中\ k=\frac{\lambda_2}{\lambda_1}. \tag{52}$$

2) $$\frac{d\eta}{d\xi}=\frac{\xi+\eta}{\xi}. \tag{53}$$

3) $$\frac{d\eta}{d\xi}=\frac{\beta\xi+\alpha\eta}{\alpha\xi-\beta\eta}. \tag{54}$$

这里 λ_1 与 λ_2 是方程

$$\begin{vmatrix} c-\lambda & d \\ a & b-\lambda \end{vmatrix}=0 \tag{55}$$

的根．如果这些根是不等的实根，则方程(50)化为(52)的形式．如果这些根是相等的，则依照 $a^2+d^2=0$ 或 $a^2+d^2\neq 0$，方程(50)分别能化为(52)或(53)的形式．如果方程(55)的根是复数，$\lambda=\alpha\pm\beta i$，则方程(51)化为(54)的形式．

分别研究方程(52)，(53)，(54)．先看下面的情形．

如果 Ox 与 Oy 两轴是互相垂直的，则一般来说，$O\xi$ 与 $O\eta$ 两轴不是互相垂直的．但为了图画的简单起见，我们仍可把它们看成是互相垂直的．此外，在变换(51)下，延 $O\xi$ 轴与 $O\eta$ 轴上的比例尺也可能变长或缩短；不一定与 Ox 轴与 Oy 轴上的比例尺一

样．但为简便起见，我们仍然可认为没有变化．因此，例如对于如图 8 中所画的同心圆形，一般则可得中心在坐标原点的类似分布的一族椭圆．

方程(52)的所有的积分线可以由关系式

$$a\eta+b|\xi|^{k}=0$$

表出，这里 a 与 b 是任意常数．

方程(52)的积分线的图形如图 10；这里，我们假定 $k>1$．在此情形下除了一条直线 $O\eta$ 外，其他所有的积分线均在坐标原点切于 $O\xi$ 轴． $0<k<1$ 的情形可以化为 $k>1$ 的情形，只要以 ξ 代 η，以 η 代 ξ，即只要 ξ 轴和 η 轴交换位置就行了．当 $k=1$ 时，方程(52)化为方程(30)，积分线分布的情形如图 7 所表示．

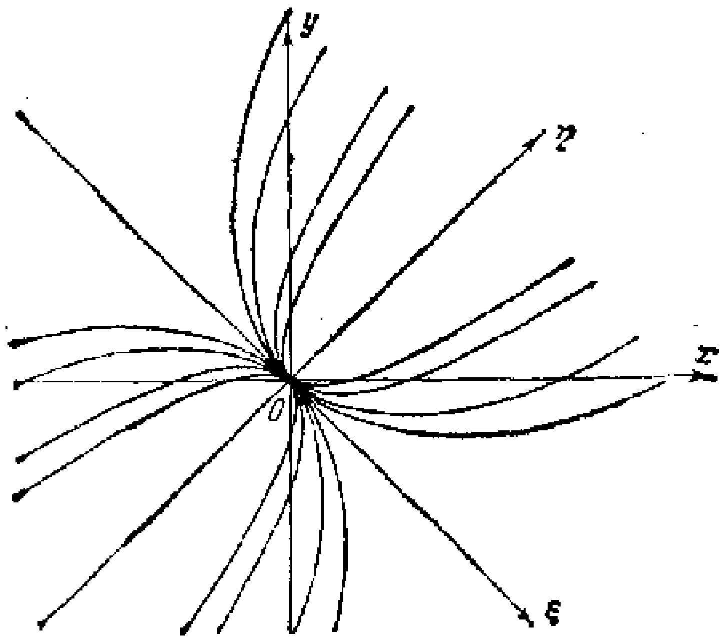


图 10

当 $k<0$ 时，方程(52)的积分线分布的情形如图 11．在此情形下，只有 $O\xi$ 轴及 $O\eta$ 轴两条积分线通过 O 点．其他的积分线向 O 点接近到一定的最小极距离之后又开始远离原点．在这种形情下，我们称这点 O 为鞍点，因为积分线类似于山隘(鞍形)的等高线的图形．

方程(53)的所有积分线可用方程

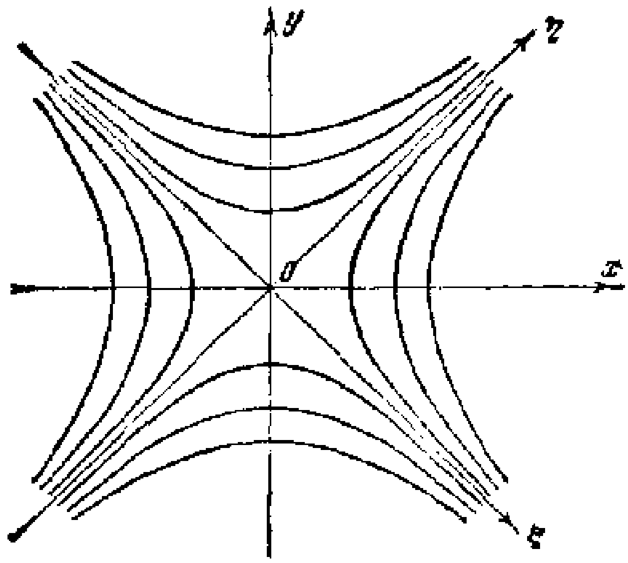


图 11

$$b\eta=\xi(a+b\ln|\xi|)$$

来表示，式中 a 与 b 为任意常数．其图形如图 12；所有的积分线在坐标原点上与 $O\eta$ 轴相切．

如果进入奇点 O 的某一邻近中的所有的积分曲线都沿一定的方向趋于 O 点，也就是在原点有确定的切线，如图 10 与 12 所表示的就是这样，则称 O 点为结点．

如果置 $\xi=\rho\cos\varphi,\ \eta=\rho\sin\varphi,$

将 $\xi,\ \eta$ 化到极坐标 $\rho,\ \varphi$，则方程(54)就很容易积分了．这方程化为

$$\frac{d\rho}{d\varphi}=k\rho,\quad k=\frac{\alpha}{\beta},$$

由此得

$$\rho=Ce^{k\varphi}. \tag{56}$$

如果 $k>0$，则所有积分曲线，当 $\varphi\to-\infty$ 时，无限旋转地趋近 O 点(图 13)．如果 $k<0$，则当 $\varphi\to+\infty$ 时，就产生同样的现象．这种 O 点叫做焦点．而当 $k=0$ 时，则积分线族(56)是由圆心在 O 的圆所组成的．一般说来，如果在 O 点的某一邻域中，都是包围着 O 点的闭曲线，其中包含 O 点，则 O 点称为中心．

如果对方程(54)的右方的分母分子加上任意高阶项，中心可

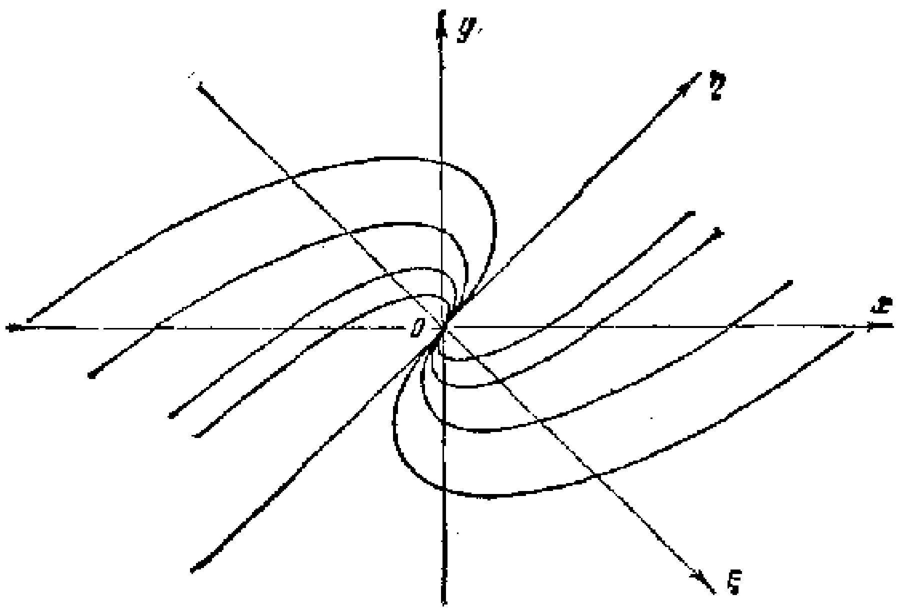


图 12

能变为焦点；因此，在中心的情形，一次项不能决定奇点附近的积分线的分布。

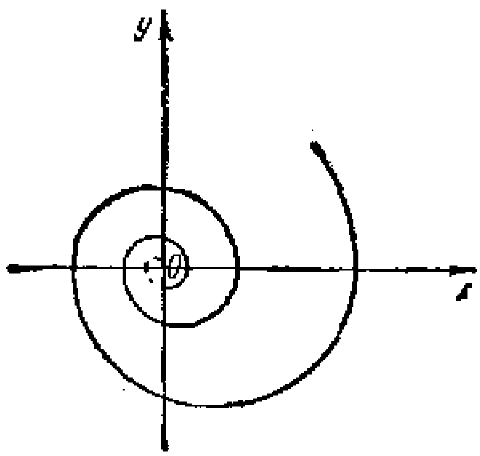


图 13

对应于方程(45)的方程(55)与示性方程(19)是一样的．因此，图10与图12表示，当数 λ_1 与 λ_2 同号时，对应于方程(6)的解

$$x=x(t),\ y=x'(t)$$

的曲线，在相平面 x, y 上的分布；图11对应于不同号的实数 λ_1 与 λ_2，而图13与图8(中心情形)对应于复数 λ_1 与 λ_2 的情况．如果 λ_1 与 λ_2 的实数部分为负的，则当 $t\to+\infty$，点 $(x(t), y(t))$ 趋于原点 O；在这种情形点 $x=0$, $y=0$ 对应于稳定的平衡情形．如果 λ_1 与 λ_2 中的一个的实数部分为正的，则在点 $x=0$, $y=0$ 上，没有稳定的平衡．

§7. 常微分方程定性理论

微分方程定性理论在常微分方程的一般理论中占有重要位

置. 由于力学及天文学的要求,在上世纪末开创了定性理论.

在很多应用问题中,要求从描述某些物理过程的微分方程,知道微分方程解的特性,并确定在自变量变动的有限或无限的变动区间中解的性质. 因此,例如在研究天体运动的天体力学中,行星或其他天体的运动是可以用微分方程来描述的,在无限增加的时间中,对于这种微分方程的解的性质的了解是很重要的.

前面我们已经说过,只有少数特别简单的方程,它的通解才能用已知函数的积分表示出来. 因此产生了直接从方程式本身,来研究微分方程的解的性质的问题. 因为微分方程的解,在平面上或空间中,可用曲线的形式来表示,因此产生了关于积分曲线的性质,其分布及其在奇点附近的结构等问题的研究. 例如,它是否分布在平面的有限部分中,是否有趋于无限远的分支,是否有闭曲线等等. 这类问题的研究构成微分方程定性理论的问题.

俄罗斯数学家李雅普诺夫及法国数学家庞卡莱是微分方程定性理论的奠基者.

在微分方程定性理论中,苏联科学所起的卓越作用,不只在于苏联学者开拓了重要的理论问题,而且还在于,在苏联最先对于解决物理与力学的问题,广泛应用了定性理论的深奥的结果.

在前一节中我们详细地研究了定性理论的一个重要问题——在奇点附近,积分曲线分布性质的问题. 现在来看定性理论的若干其他基本问题.

稳定性 在本章开始所研究的例子中,系统的平衡态的稳定性与不稳定性的问题,不是由微分方程的研究来解决,而是基于物理情况来解决的. 因此,在例 3 中很显然,如果处于平衡态 OA 的摆用外力移到某一附近位置 OA',即是初始条件很小的变动,则摆的此后的运动不能远离平衡位置,初始偏差 OA' 越小,则摆的离平衡位置的偏差也越小,即是在这种情形平衡位置是稳定的.

在其他更复杂的问题中,平衡态稳定性问题的解决很复杂,而且只能借助于对应的微分方程的研究来解决. 关于运动稳定性的问题与关于平衡态稳定性的问题是密切相关的. 在这个领域内奠

基的结果是属于李雅普诺夫的.

设某个物理过程用方程组

$$\frac{dx}{dt}=f_1(x,\ y,\ t),\qquad \frac{dy}{dt}=f_2(x,\ y,\ t) \tag{57}$$

来描述. 为了简便起见,我们只研究两个微分方程的方程组,虽然我们以后的结论,对于具多个微分方程的方程组也是适用的.方程(57)的每一个特解由两个函数 $x(t)$ 与 $y(t)$ 所组成,在本节中我们依照李雅普诺夫的说法,称它为**运动**. 假设 $f_1(x,\ y,\ t)$ 与 $f_2(x, y, t)$ 有连续偏微商. 已经证明, 如果在任何时间 $t=t_0$ 给定函数值 $x(t_0)=x_0$ 与 $y(t_0)=y_0$,则微分方程组(57)的解是唯一决定的.

满足初始条件

$$x=x_0 \quad 与 \quad y=y_0 \quad 当 \quad t=t_0$$

的方程组(57)的解用 $x(t,\ x_0,\ y_0)$, $y(t,\ x_0,\ y_0)$ 表示.

解 $x(t,\ x_0,\ y_0)$, $y(t,\ x_0,\ y_0)$ 称为**按李雅普诺夫的意义是稳定的**, 如果对所有 $t>t_0$ 函数 $x(t,\ x_0,\ y_0)$, $y(t,\ x_0,\ y_0)$ 变化得很少,只要初始值 x_0 与 y_0 变动得很小.

更确切地说:对于按李雅普诺夫的意义是稳定的解,则差数

$$\begin{gathered}|x(t,\ x_0+\delta_1,\ y_0+\delta_2)-x(t,\ x_0,\ y_0)|,\\ |y(t,\ x_0+\delta_1,\ y_0+\delta_2)-y(t,\ x_0,\ y_0)|\end{gathered} \tag{58}$$

对所有的时间 $t>t_0$, 可能小于任何预先给定的正数 ε, 只要把 δ_1 与 δ_2 的绝对值取得足够小就行了.

所有不是按李雅普诺夫意义稳定的运动都称为不稳定.

把所要研究的运动 $x(t,\ x_0,\ y_0)$, $y(t,\ x_0,\ y_0)$ 李雅普诺夫称之为未被扰动的,而具有很接近的初始条件的运动 $x(t,\ x_0+\delta_1,\ y_0+\delta_2)$, $y(t,\ x_0+\delta_1,\ y_0+\delta_2)$ 称为被扰动的. 因此,未被扰动的运动按李雅普诺夫意义稳定, 即表示对所有 $t>t_0$ 被扰动的运动与未被扰动的运动相差很小.

平衡的稳定性是运动稳定性的特例,在此情形下,对应的未被

扰动的运动是

$$x(t,\ x_0,\ y_0)\equiv 0 \quad 与 \quad y(t,\ x_0,\ y_0)\equiv 0.$$

反之，系统(57)的任何运动 $x=\varphi_1(t)$ 与 $y=\varphi_2(t)$ 的稳定性问题可以化为某个微分方程组的平衡的稳定性问题．为此，只要把方程组(57)中，旧的未知函数 $x(t)$ 与 $y(t)$ 换为新的未知函数

$$\xi=x-\varphi_1(t),\ \eta=y-\varphi_2(t). \tag{59}$$

在这种变换下方程式(57)的运动 $x=\varphi_1(t)$, $y=\varphi_2(t)$ 与运动 $\xi\equiv 0$ 与 $\eta\equiv 0$ 即平衡态相对应．以下我们假定已经用过了变换(59)，只对解 $x\equiv 0$ 与 $y\equiv 0$ 来研究按李雅普诺夫意义的稳定性．

现在按李雅普诺夫意义的稳定性的条件就表示，如果 δ_1 与 δ_2 足够小，则被扰动的运动在平面$(x,\ y)$上的轨线，对任何 $t>t_0$，不走出以点 $x=0$, $y=0$ 为心，边长为 2ε，边平行于坐标轴的正方形．

我们有兴趣于以下的情形，我们并不能求出方程组(57)的积分，然而仍能作出有关未被扰动的运动的稳定性或不稳定性的结论．

在实际上很重要的关于炮弹运动和飞机运动的稳定性问题，在天体力学中重要的关于行星和其他天体运动的轨道的稳定性问题，都归结为这类的研究．

假设，函数 $f_1(x,\ y,\ t)$ 与 $f_2(x,\ y,\ t)$ 可以写成以下的形式

$$\begin{aligned} f_1(x,\ y,\ t)&=a_{11}x+a_{12}y+R_1(x,\ y,\ t),\\ f_2(x,\ y,\ t)&=a_{21}x+a_{22}y+R_2(x,\ y,\ t), \end{aligned} \tag{60}$$

其中 a_{ij} 是常数，而 $R_1(x,\ y,\ t)$ 与 $R_2(x,\ y,\ t)$ 是 x, y 与 t 的函数，适合

$$|R_1(x,\ y,\ t)|\leqslant M(x^2+y^2) \quad 与 \quad |R_2(x,\ y,\ t)|\leqslant M(x^2+y^2), \tag{61}$$

其中 M 为一正常数．

设在(60)所表示的方程组(57)中，把 $R_1(x,y,t)$ 与 $R_2(x,y,t)$ 抛弃，则得到常系数的微分方程组

$$\begin{aligned}\frac{dx}{dt}&=a_{11}x+a_{12}y,\\ \frac{dy}{dt}&=a_{21}x+a_{22}y,\end{aligned}\tag{62}$$

把它称作非线性方程组(57)的第一次近似方程组。

在李雅普诺夫以前，在稳定性问题的研究上，基本上，限制于第一次近似的稳定性的研究，即是对方程式(62)的稳定性的研究，把得到的结果看作适用于基本的非线性系统(57)关于稳定性问题的解答．李雅普诺夫第一个证明，在一般的情形下，这个结论不成立．另一方面，他给出了一系列非常广泛的条件，当这类条件满足时，非线性方程式关于稳定性的问题由第一次近似完全解决．这类条件中之一如下，如果方程

$$\begin{vmatrix}a_{11}-\lambda & a_{12}\\ a_{21} & a_{22}-\lambda\end{vmatrix}=0$$

的所有的根 λ 的实数部分为负的，而且函数 $R_1(x, y, t)$ 与 $R_2(x, y, t)$ 满足条件(61)，则解 $x(t)\equiv0$, $y(t)\equiv0$ 按李雅普诺夫意义是稳定的．如果至少有一根 λ 其实数部分为正的，则当满足条件(61)时，解 $x(t)\equiv0$，$y(t)\equiv0$ 是不稳定的．李雅普诺夫也给了运动稳定性与不稳定性的一系列的其他充分条件[1]．苏联数学家继续了李雅普诺夫的工作，并取得了极大的成就．

如果方程(57)的右边与 t 无关，则用方程组(57)的第二个方程除第一个方程，我们得到

$$\frac{dy}{dx}=\frac{f_1(x, y)}{f_2(x, y)}.\tag{63}$$

坐标原点对这个方程是奇点．在平衡态是稳定的情形，在这点可能是焦点，结点或中心，但不能是鞍点．

因此，根据奇点的性质可以来判断平衡态的稳定性或不稳定性．

积分曲线的全局分布 有时在微分方程组的定义区域上，“全局”地作出积分曲线的分布图是很重要的，这里不必考虑保持比例

1) 李雅普诺夫运动稳定性通论．

尺的大小．这方程组所定义的向量场的空间将看作某个物理过程的相空间．则对应的微分方程组的积分曲线的图可以给我们，对这方程式所可能产生的所有的过程（运动）的特性，提供了一些概念．在图 10 到图 13 我们在孤立奇点的邻域内，对积分曲线的分布，作出了详细的图形．

微分方程理论的最基本的问题之一是用尽可能简单的方法来找出在微分方程所有的定义区域中积分曲线族的分布图形，就是这个微分方程组的积分曲线的"全局"分布的研究．这个问题对于维数大于 2 的空间差不多完全没有研究过．甚至于在 $M(x, y)$ 与 $N(x, y)$ 是多项式的情形下，对一个如下形式

$$\frac{dy}{dx}=\frac{M(x, y)}{N(x, y)} \tag{64}$$

的方程而言，这问题的解决也还差得很远．

以下我们将假设，函数 $M(x, y)$ 与 $N(x, y)$ 有一阶连续偏微商．

如果微分方程(64)的右边在一单连通的区域 G 中已给定，又 G 中所有点都是常点，则积分曲线族可以用平行直线段族的图形来表示，因为这时 G 中每点均有一积分线经过，并且任何两积分线不相交．而对于更一般的情形的方程(64)，它可能有奇点，积分曲线的结构可能很复杂．方程(64)有无限多个奇点(即分子分母同时变成零的点)的情形，至少当 $M(x, y)$ 与 $N(x, y)$ 是多项式时，可能看作是例外．因此，我们只限于研究方程(64)有有限多个孤立奇点的这类情形．要作出这个方程的所有积分线的分布图，则靠近每一个奇点的积分线的分布是很关键的．

所谓的**极限环**对于方程(64)的所有积分线的分布图，也是非常本质的．我们来研究一下方程

$$\frac{d\rho}{d\varphi}=\rho-1, \tag{65}$$

式中 ρ 与 φ 是在平面 (x, y) 上的极坐标．

方程(65)的所有积分线的全体可以用公式

$$\rho = 1 + Ce^{\varphi} \tag{66}$$

来表示，式中 C 是任意常数，对于不同的积分线取不同的值．如果 $C<0$，为了要使 ρ 不为负数，必须 φ 不取大于 $-\ln|C|$ 的值．积分线族可分为

1) 圆 $\rho=1(C=0)$；

2) 由坐标原点走出的螺线，它由内部趋近这个圆，当 $\varphi \to -\infty(C<0)$ 时；

3) 由圆外向圆趋近的螺线，当 $\rho \to -\infty(C>0)$ 时（图 14）．

圆 $\rho=1$ 称为方程(65)的极限环．一般来说，闭积分线 l 称为**极限环**，如果它可能被包含在一环形域内，这个环形域中所有的点对微分方程来说都是常点，又在这环形域中的其他积分曲线都不是闭的．

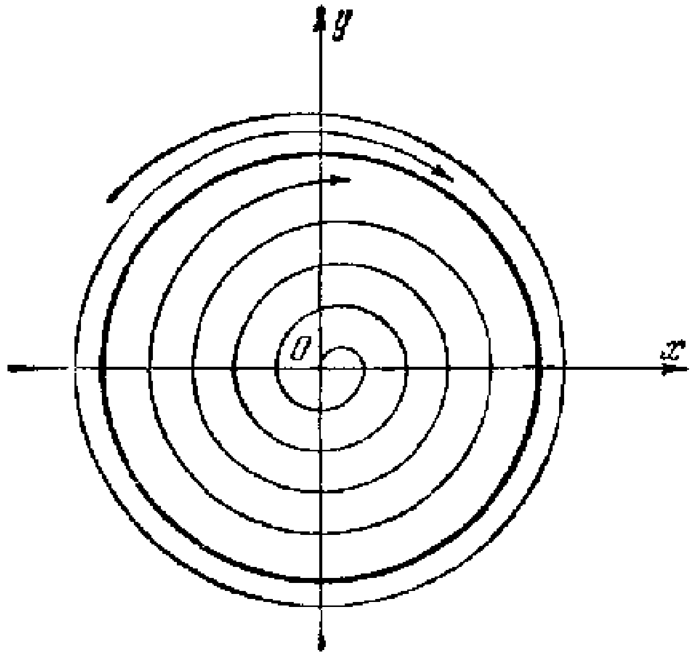


图 14

由方程(65)可见，圆上所有的点对它都是常点．这表示极限环上的小段弧线与所有其他积分线上的小段弧没有什么区别．

在相平面(x, y)上，每个闭积分曲线对应于方程组

$$\frac{dx}{dt} = N(x, y), \quad \frac{dy}{dt} = M(x, y) \tag{67}$$

的一个周期解$[x(t), y(t)]$，而这方程组描述某个物理系统的变化规律．当 $t\to+\infty$ 时在相平面上趋近于极限环的这种积分线对应于当 $t\to+\infty$时趋近于周期解的运动．

设对任何充分接近极限环 l 的点(x_0, y_0)，取作 $t=t_0$ 时的初始点，对于方程组(67)，由点$(x(t), y(t))$所描述的对应的积分线，当 $t\to+\infty$ 时，趋近于(x, y)平面上的极限环 l（这表明，所描述的运动趋近于周期解）．在这种情形对应的极限环称为**稳定的**．对应于这个极限环的振动相当于物理中的自振．在某些自振系统中，可能存在几个驻定的、但是不同振幅的振动过程．那一个出现是由初始条件来决定．如果在这个自振系统中出现的过程是可以用(67)形式的方程描述的话，则在相平面上这种自振系统对应于几个极限环．

到现在为止，即使对于已知的微分方程，求极限环近似方法的问题也还没有得到满意的解答．解决这个问题最广泛用到的方法是庞卡莱所建议的作“无切点的环线”的方法．它基于下面的定理的．设在(x, y)平面上已找到两个这样的闭线 L_1 与 L_2(环线)，它们具有下列性质：

1) L_2 线在 L_1 所包的区域的内部；

2) 在 L_1 与 L_2 之间的环域 Ω 中，方程(64)没有奇点；

3) L_1 与 L_2 上每点都有切线，而且这些切线的方向与方程(64)在这点的向量场的方向不相重合；

4) 在 L_1 及 L_2 的任一点上，对环域 Ω 的内法线和分量为$[N(x, y), M(x, y)]$的向量之间的角度的余弦都是同一符号的．

庞卡莱把 L_1 与 L_2 称为**无切点的环线**．在 L_1 与 L_2 之间至少有一条方程(64)的极限环．这个定理的证明是基于下面的很显然的事实．设当 t 增加时(或当 t 减少时)，方程(64)的所有积分线

$$x=x(t),\quad y=y(t)$$

(同样，可以是方程(67)的积分线，其中 t 为参数)都穿过 L_1 或 L_2 进入 L_1 与 L_2 之间的环域 Ω．它们必然要旋转接近于某个闭线

l_1, l_1 在 L_1 与 L_2 之间，这是因为这个环域中的任一积分线均不能走出去，而在环域内部又没有奇点。

寻求无切环线本身便是相当复杂的问题．还不知道任何一般的方法．对个别特例则找到了无切环线，由此证明了极限环的存在。

在无线电技术中对于真空管发动器方程(16)找出极限环(自振过程)具有很大的意义．对形式(16)的方程，大约在20年前，克雷洛夫与波果留博夫对于有极限环的情况，提出近似计算的方法．大约也在这时候，苏联物理学家曼德尔斯塔姆巴巴勒克西与安德洛诺夫证明了，为了这个目的，可以利用所谓小参数法，虽然这个方法在实际上过去已应用过了，但是运用这规律的严格理论基础则是没有过的．为了分析自振系统方程，安德洛诺夫第一个开始系统地应用了李雅普诺夫与庞卡莱以前在微分方程中所发展的方法．这样，他得到了一系列重要的结果。

如前面已说过的，在物理中"粗糙"的方程组(参看§3)有很大的作用．安德洛诺夫与邦特里雅金作出了一系列的典型的研究，对(64)形式的粗糙的微分方程在(x, y)平面上的积分线的分布图都可以由这些典型图合并而成．很早便知道，当方程(64)遇微小变化时，中心很容易就被破坏了．因此，如果方程组是粗糙的，在方程(64)的积分线的分布图中，不能出现中心(即包含奇点在内的一族闭积分线)。

积分曲线全局分布的问题还远没有完全解决．我们注意到，同样可以想到更简单的问题，实代数曲线，在平面上可能有那些类型，即是由方程

$$P(x, y)=0$$

所定的曲线的类型，其中 $P(x, y)$——n 次多项式，这问题也是还远没有完全解决．只有当 $n<6$ 时，这些曲线有什么样的类型才完全知道．

方程(64)的解决定在平面上的运动．如果对平面上每一点(x_0, y_0)，作对应点$[x(t, x_0, y_0), y(t, x_0, y_0)]$，其中 $x(t, x_0, y_0)$，

$y(t, x_0, y_0)$为方程组(64)，当 $t=t_0$ 时，具初值 $x=x_0$ 与 $y=y_0$ 的解，则得到以 t 为参数的平面上的点变换．与参数有关的类似的变换及其所产生的运动可以在球面、环面及其他流形上进行研究．在动力系统理论中，研究这些运动的性质．这些运动在每一点的邻域中，也是某个微分方程组的解．在最近十年中动力系统的理论，在苏联数学家斯捷巴诺夫、辛钦、波果留博夫、克雷洛夫、马尔科夫、涅梅茨基及其他人，以及比尔可夫及其他外国学者的工作中，得到广泛的发展．

在这一章中我们对常微分方程理论的现代情况给了一个概述，并力求说明这个理论中所研究的问题．我们的叙述从任何角度也不能追求全面．我们被迫抛弃了方程理论的许多部分的讨论，它们或者是更特殊问题的研究，或者它们要求读者具备比我们假定所具备的数学知识要有更多的知识．例如我们一点没有谈到一个广泛而重要的部门，具有复变数的微分方程理论的研究．我们没有可能谈到所谓边界值问题，特别是在应用中，有很大意义的固有函数理论．

我们只能很少注意到，微分方程数值近似解及解析解的方法等等．要想知道这些问题我们只有向读者推荐去读专门的书籍．

文　献

大学课本

И. Г. 彼得罗夫斯基，常微分方程论讲义，商务印书馆，1953年版．

В. В. 斯捷班诺夫，常微分方程，商务印书馆，1953年版．

Л. Э. Эльсгольц, Обыкновенные дифференциальные уравнения, Гостехиздат, 1954.

这本书包括了微分方程理论的基本要点，也叙述了微分方程积分的近似方法．本书的对象主要是工程师及工科大学生．

单行本

А. А. Андронов и С. Э. Хайкин, Теория колебаний, ч. 1, ОНТИ, 1937.

本书中研究应用于振动理论中的微分方程．在研究这些问题的时候，应用了微分

方程定性理论的方法。

B. B. 涅梅茨基与 B. B. 斯捷班诺夫，微分方程定性理论，科学出版社，1956 年版。

书中叙述了这个理论的基础。本书的对象是大学学生及科学工作者。

秦元勋 译

周毓麟 校

第六章 偏微分方程

§1. 绪 论

在研究自然现象的时候，我们遇到偏微分方程的机会并不比遇到常微分方程的机会少．照例，当我们所研究的自然现象是用多变量的函数描述的时候，我们便遇到偏微分方程．在自然的研究中，产生了一类偏微分方程，这类偏微分方程是现在研究得最多的，或许是人类知识系统中最重要的，这就是数学物理方程．

首先我们来考虑任何介质的振动．设这介质的每点在平衡状态时的位置是 x, y, z，在振动时这点在时刻 t 的位移是一个向量 $\boldsymbol{u}(x, y, z, t)$，这向量与点的原来的位置 x, y, z 以及时间 t 有关．于是我们所要研究的现象就可用向量场来描写．但是不难看到，即使知道了这个向量场——介质的每一点的位移场——仍不足以完全了解这介质的振动．譬如我们还需要知道介质的每一点的密度 $\rho(x, y, z, t)$，每点的温度 $T(x, y, z, t)$ 以及内力，所谓内力就是这物体的每一部分所受的其余部分的作用力．

在空间和时间过程中，进行的物理过程，物理现象总是归结为空间某区域的点上物理量随着时间的变化．正如我们在第二章中（第一卷）所看到的，这些物理量是以四个自变数 x, y, z, t 的函数来描写的，这里 x, y, z 是空间任一点的坐标，t 是时间．

大家知道，物理量可以有不同的性质．其中一些是用数值来完全刻划的，如温度、密度等等，这些量叫做纯量．另一些是有方向的，那便是向量，如速度、加速度、电场的引力等等．向量不仅可以用它的长度和方向来表示，而且可用它的"分量"来表示，也就是把它分为三个互相垂直的向量的和，譬如这三个向量可以平行于坐标轴．

在数学物理中，纯量或纯量场可以用一个含有四个自变数的函数来表示，而在整个空间定义的向量，即通常所谓向量场，则可用三个含有四个自变数的函数来表示。这向量可以写成以下的形式

$$\boldsymbol{u}(x, y, z, t),$$

这里黑体字表示，$\boldsymbol{u}$ 是向量，或写成

$$u_x(x, y, z, t), u_y(x, y, z, t), u_z(x, y, z, t),$$

其中 u_x, u_y, u_z 表示向量在坐标轴上的投影。

除向量与纯量以外，在物理中我们还遇到更复杂的量，如物体在一点的张力状态。这样的一些量叫做张量，但若选定坐标以后，每个这样的量仍然可以用一些含有四个自变数的函数来表示。

所以所有可能的物理现象总可以用某些多变数函数来描述。当然决不能强求这种描述是绝对精确的。

譬如，当用四个自变数的函数来描述介质的密度时，我们撇开了这样一件事实，即"在一点的"密度实际上是不存在的。我们所研究的物体具有分子构造，而且这些分子并不相互紧密地接在一起，而相互有一定的距离。它们之间的距离比起分子本身的大小来，一般是很大的。因此我们所说的密度是包含在某个小的——但不是过分小的——体积中的物质的质量对这体积的比。我们了解在一点的密度的意思是这个比值当体积缩小时的极限。介质的温度也是一个极抽象而且理想化了的概念。物体的温度是这物体的分子作无秩序运动所引起的。分子的能量是不同的，但是如果我们考虑包含着许多分子的一个体积，那末这些分子的无秩序运动的平均的能量便确定了这样一种特征，我们称之为温度。

与此相似，当我们讲到气体或液体在器壁上的压力时，我们不应该把这种压力想象为似乎有一部分气体或液体真正地压在器壁上。事实上，这一小部分在作无秩序运动时碰撞器壁又由器壁弹回。我们所描述为器壁上的压力者实际上是器壁的某部分所受到的大量分子的衡量的总和，这部分器壁从我们的观点来说是小的，但与气体或液体的分子之间的距离比较起来仍是很大的。这种类

似的例子很多．数学物理所考虑的都是理想化了的量，它们都是从这些量的许多具体性质而且考虑它们的平均数值而抽象得来的．

这种理想化似乎是粗糙的，但是我们即将看到，它给我们带来了很大的方便，它使我们能很好地分析许多复杂的现象，从其中提出主要方面，撇开次要的东西．

数学物理方程的研究对象仍是这些经理想化了的现象之间的联系，而这些现象是用一些多变数函数来描述的．

§2．最简单的数学物理方程

物理量之间的基本的联系和关系被表示为物理和力学的定律．这种联系是多种多样的，并且从这些联系可以导出更多更复杂的联系，这就是那些定律的数学推论．物理和力学的定律可用数学的语言写成偏微分方程的形式，有时也可以写成积分方程的形式，这些方程便是未知函数之间的联系．为了了解问题的实质，我们列举数学物理方程的一些范例：

试把规定介质点的运动的一些基本物理定律，写成数学的表现形式．

质量守恒方程和热能守恒方程 **1**．首先想象在空间中取出一个固定的体积 Ω，我们来描述在每个这样的体积中的质量的守恒定律．为此需计算这体积中的物质的质量．这质量 $M_\Omega(t)$ 可用以下的积分来表示：

$$M_\Omega(t)=\iiint\limits_\Omega \rho(x,\ y,\ z,\ t)\,dx\,dy\,dz.$$

当然这质量不一定是常数；因为在振动的过程中，有一部分物质进入这个体积，也有一部分物质移出这个体积，所以每一点的密度在运动过程中是变化的．将上面的积分对时间 t 求微分，我们就得到质量的变化速度，

$$\frac{dM_\Omega}{dt}=\iiint\limits_{\Omega}\frac{\partial\rho}{\partial t}dx\,dy\,dz.$$

这个质量的变化速度也可以用另一种方法来计算. 设曲面 S 是我们的体积 Ω 的边界, 我们可以表述每秒钟流过这边界面 S 的物质的量, 从 Ω 中流出的物质的量须冠以负号. 为此我们考虑边界面 S 的一小部分 ds, 这部分取得充分小, 可以把它看做是平面而且它上面的质点的位移可以看做是一致的. 从时间 t 到 $t+\Delta t$ 的过程中, 我们考察这部分面积上的质点. 我们首先注意

$$\boldsymbol{v}=\frac{d\boldsymbol{u}}{dt}$$

是每部分的速度. 经过时间 dt 以后, ds 上的质点移动了 $\boldsymbol{v}dt$, 而移到 ds_1 的位置, 且原来在 ds_2 上的质点则移动到 ds 上来(图 1).

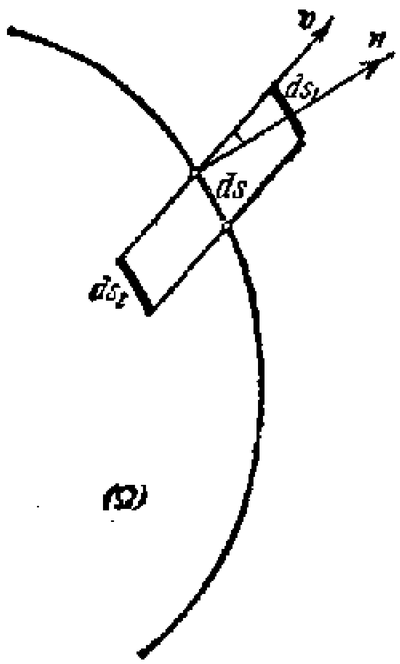


图 1

因此, 在这段时间中, 原来在 Ω 内的整个柱体 ds_2ds 的物质从 Ω 中流出, 且结果流到 $ds\,ds_1$ 的位置. 这柱体的高等于 $v\,dt\cdot\cos(\boldsymbol{n},\ \boldsymbol{v})$, 其中 $\boldsymbol{n}$ 表示曲面的外法线方向; 柱体的体积等于

$$v\cos(\boldsymbol{n},\ \boldsymbol{v})ds\,dt,$$

而质量等于

$$\rho v\cos(\boldsymbol{n},\ \boldsymbol{v})ds\,dt.$$

把所有这样的柱体加起来, 我们就得到在时间 dt 从 Ω 中流出的质量的表达式

$$\iint\limits_{S}\rho v\cos(\boldsymbol{n},\ \boldsymbol{v})ds\,dt.$$

这里, 在速度是向着 Ω 内的地方, 余弦的符号是负的, 这就是说, 在积分中流入的物质算带有负号. 介质的运动速度和其密度的乘积称为流量. 质量的流量是一向量 $\boldsymbol{q}=\rho\boldsymbol{v}$.

为了得到物质流出的速度, 只须将以上得到的表达式除以 dt,

所以我们得物质流出的速度的表达式

$$\iint_s \rho v_n\, ds = \iint_s q_n\, ds,$$

式中 $v_n = v\cos(\boldsymbol{n},\ \boldsymbol{v})$； $q_n = q\cos(\boldsymbol{n},\ \boldsymbol{q})$.

向量 $\boldsymbol{v}$ 沿法线方向的分量可以用 $\boldsymbol{v}$ 和 $\boldsymbol{n}$ 沿坐标轴的分量来表示. 由解析几何可以知道

$$v_n = v\cos(\boldsymbol{n},\ \boldsymbol{v}) = v_x\cos(\boldsymbol{n},\ \boldsymbol{x}) + v_y\cos(\boldsymbol{n},\ \boldsymbol{y}) + v_z\cos(\boldsymbol{n},\ \boldsymbol{z}),$$

因此我们可以把物质流出的速度的表达式改写成

$$\iint_s \rho(v_x\cos(\boldsymbol{n},\ \boldsymbol{x}) + v_y\cos(\boldsymbol{n},\ \boldsymbol{y}) + v_z\cos(\boldsymbol{n},\ \boldsymbol{z}))\, ds.$$

因为只有在质量经过曲面 S 流入或流出时, Ω 中的质量才能发生变化, 根据质量守恒定律, 以上计算质量变化的两种方法应该给出相同的结果.

因此, 命 Ω 中质量的变化速度等于其输入的速度, 乃得

$$\iiint_\Omega \frac{\partial\rho}{\partial t}\, dx\, dy\, dz$$

$$= -\iint_s (\rho v_x\cos(\boldsymbol{n},\ \boldsymbol{x}) + \rho v_y\cos(\boldsymbol{n},\ \boldsymbol{y}) + \rho v_z\cos(\boldsymbol{n},\ \boldsymbol{z}))\, ds$$

$$= -\iint_s (q_x\cos(\boldsymbol{n},\ \boldsymbol{x}) + q_y\cos(\boldsymbol{n},\ \boldsymbol{y}) + q_z\cos(\boldsymbol{n},\ \boldsymbol{z}))\, ds.$$

此关系式对任何体积 Ω 皆成立. 这就叫做**连续性方程**.

这等式右边的积分可以用奥斯特洛格拉特斯基公式化为体积分. 此公式在第二章(第一卷)已引入, 由这公式得

$$\iint_s (\rho v_x\cos(\boldsymbol{n},\ \boldsymbol{x}) + \rho v_y\cos(\boldsymbol{n},\ \boldsymbol{y}) + \rho v_z\cos(\boldsymbol{n},\ \boldsymbol{z}))\, ds$$

$$= \iiint_\Omega \left(\frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z}\right) d\Omega.$$

由此可得

$$\iiint_\Omega \left(\frac{\partial\rho}{\partial t} + \frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z}\right) d\Omega = 0.$$

我们得到以下的结果：在任何体积 Ω 上，函数

$$\frac{\partial\rho}{\partial t}+\frac{\partial(\rho v_x)}{\partial x}+\frac{\partial(\rho v_y)}{\partial y}+\frac{\partial(\rho v_z)}{\partial z}$$

或

$$\frac{\partial\rho}{\partial t}+\frac{\partial q_x}{\partial x}+\frac{\partial q_y}{\partial y}+\frac{\partial q_z}{\partial z}$$

的积分为零，故这函数恒等于零．这样一来，我们得到连续性方程的另一形式——微分形式

$$\frac{\partial\rho}{\partial t}+\frac{\partial(\rho v_x)}{\partial x}+\frac{\partial(\rho v_y)}{\partial y}+\frac{\partial(\rho v_z)}{\partial z}=0. \qquad (1)$$

方程(1)是用偏微分方程的语言来叙述物理的定律的典型的例子．

2．现在我们再来考虑一个这种问题，即热量传导的问题．

如果在整个介质中，有热量从一点传播到另一点．我们想象在指定的介质内部取出某一曲面，那么在热量的传播过程中，将有热量通过曲面的每一小部分 ds．我们所考虑的热的传播现象可以用一特殊的向量——"热流向量"来描述，我们用 τ 来表示这热流向量．于是在每秒钟通过某面积 ds 的热量有表达式 $\tau_n ds$，这表达式和我们以上推导连续性方程时所得到的流体在每秒钟通过面积 ds 的流量的表达式 $q_n ds$ 相类似．这里用热流向量 τ 来代替以前的流体的流量 $\boldsymbol{q}=\rho\boldsymbol{v}$．

连续性方程表示流体在运动中的质量守恒定律，用推导连续性方程的同样的方法，我们可以得到一个新的偏微分方程，它表示能量的守恒定律．

在每一点热能的体密度 Q 可以用下列公式来表示：

$$Q=CT$$

其中 C 是体热容，T 是热度．

于是不难得到方程

$$C\frac{\partial T}{\partial t}+\frac{\partial\tau_x}{\partial x}+\frac{\partial\tau_y}{\partial y}+\frac{\partial\tau_z}{\partial z}=0. \qquad (2)$$

如果把"密度"换为"能量密度"，"流量"换为"热流向量"，则此方程的推导和连续性方程的推导完全一样．于此我们须设：在我们考

虑热导现象的介质中，到处都没有热能产生．但如果在介质中有热量输出，则方程(2)就应该有所变化．设 q 是热能"输出的密度"，即在单位时间中，单位体积的介质所输出的热能，则热能的守恒方程就具有下列比较复杂的形式

$$C\frac{\partial T}{\partial t}+\frac{\partial \tau_x}{\partial x}+\frac{\partial \tau_y}{\partial y}+\frac{\partial \tau_z}{\partial z}=q. \tag{3}$$

3．将方程(1)对时间求微分，我们还可以得到一个和连续性方程(1)同类型的方程．我们假设气体围绕着它的平衡位置作微小振动．于是我们可认为在这种振动过程当中，密度的变化是很小的，即 $\frac{\partial \rho}{\partial t}$, $\frac{\partial \rho}{\partial x}$, $\frac{\partial \rho}{\partial y}$ 和 $\frac{\partial \rho}{\partial z}$ 都是很小的量 且其与 v_x, v_y, v_z 的乘积可略去．于是

$$\frac{\partial \rho}{\partial t}+\rho\left(\frac{\partial v_x}{\partial x}+\frac{\partial v_y}{\partial y}+\frac{\partial v_z}{\partial z}\right)=0.$$

将这方程对时间求微分，且略去 $\frac{\partial \rho}{\partial t}$ 和 $\frac{\partial v_x}{\partial x}$, $\frac{\partial v_y}{\partial y}$, $\frac{\partial v_z}{\partial z}$ 的乘积，则得

$$\frac{\partial^2 \rho}{\partial t^2}+\rho\left(\frac{\partial\left(\frac{dv_x}{dt}\right)}{\partial x}+\frac{\partial\left(\frac{dv_y}{dt}\right)}{\partial y}+\frac{\partial\left(\frac{dv_z}{dt}\right)}{\partial z}\right)=0. \tag{4}$$

运动方程　1．介质的运动方程或平衡方程是用微分方程表示物理定律的重要范例．设介质是由许多物质的部分所组成的，并且这些部分以某速度运动着．和第一个例子一样，想象在我们的介质中取出某个体积 Ω，在 Ω 中含有介质的一部分物质，命 Ω 的边界是 S，对于 Ω 中的这部分物质，写出牛顿的第二定律．牛顿第二定律是说：在介质的整个运动过程，某系统所有点的总动量的变化速度等于作用在这体积上的所有力的冲量之和．从力学中知道，动量是向量

$$\boldsymbol{P}=\iiint_{\Omega}\rho \boldsymbol{v}\, d\Omega.$$

原来以密度 ρ 填满小体积 $d\Omega$ 的那一部分物质，经时间 Δt 以后，将以新的密度 ρ' 填满新的体积 $d\Omega'$，但其质量不变

$$\rho' d\Omega' = \rho d\Omega.$$

在这一段时间中，速度由 $\boldsymbol{v}$ 变到 $\boldsymbol{v}'$，这种速度的变化 $\Delta\boldsymbol{v}=\boldsymbol{v}'-\boldsymbol{v}$ 引起了动量的变化

$$\rho'\boldsymbol{v}' d\Omega' - \rho\boldsymbol{v} d\Omega = \rho\boldsymbol{v}' d\Omega - \rho\boldsymbol{v} d\Omega = \rho\Delta\boldsymbol{v} d\Omega.$$

于是在单位时间中，动量的变化为

$$\rho \frac{\Delta\boldsymbol{v}}{\Delta t} d\Omega \approx \rho \frac{d\boldsymbol{v}}{dt} d\Omega.$$

把 Ω 中所有的部分加起来，我们就得出它的动量的变化速度

$$\iiint_{\Omega} \rho \frac{d\boldsymbol{v}}{dt} d\Omega$$

或 $$\iiint_{\Omega} \rho \frac{dv_x}{dt} d\Omega, \quad \iiint_{\Omega} \rho \frac{dv_y}{dt} d\Omega, \quad \iiint_{\Omega} \rho \frac{dv_z}{dt} d\Omega.$$

$\Big($这里的微商 $\frac{dv_x}{dt}$, $\frac{dv_y}{dt}$, $\frac{dv_z}{dt}$ 并不是在空间某一点的而是在已知部分上 $\boldsymbol{v}$ 的各分量的变化速度，我们不用符号 $\frac{\partial}{\partial t}$ 而用 $\frac{d}{dt}$，就是为了表明这一点．显然 $\frac{d}{dt}=\frac{\partial}{\partial t}+v_x\frac{\partial}{\partial x}+v_y\frac{\partial}{\partial y}+v_z\frac{\partial}{\partial z}$.$\Big)$

作用在某体积上的力可能有两类：一类是作用在这体积中的每一部分上的力；另一类是作用在这体积的表面 S 上的力．第一类是远作用力，而第二类是近作用力．

为了阐明上述的概念，假设我们所考虑的介质是液体．于是作用在面积元素 ds 上的表面力是 $p ds$，其中 p 是液体的压力，而此力的方向和外法线相反．

若以 $\boldsymbol{n}$ 表示沿曲面 S 的外法线的单位向量，则作用在面积元素 ds 上的力将等于

$$-p\boldsymbol{n}\, ds.$$

此外，设 $\boldsymbol{F}$ 是作用在单位体积上的外力．于是我们的方程可以写成

$$\iiint_{\Omega} \rho \frac{d\boldsymbol{v}}{dt} d\Omega = \iiint_{\Omega} \boldsymbol{F} d\Omega - \iint_{S} p\boldsymbol{n}\, ds.$$

这就是运动方程的积分形式，就象连续性方程可化为方程(1)一样，这方程也可以化为微分方程的形式．结果我们得方程组

$$\rho\frac{dv_x}{dt}+\frac{\partial p}{\partial x}=F_x,\ \rho\frac{dv_y}{dt}+\frac{\partial p}{\partial y}=F_y,\ \rho\frac{dv_z}{dt}+\frac{\partial p}{\partial z}=F_z. \tag{5}$$

这方程组是牛顿的第二定律的微分表示形式．

2．弦振动方程是用微分方程表示力学定律的另一范例．所谓弦是指弹性物质构成的细而长的物体，由于它很细，所以可认为是极其柔软的．在考虑弦的振动时，总假设它是被拉紧着的．想象在弦上的每一点 x 作一断面，则这断面将弦分为两边，且由每一边有一力作用在另一边上，此力的大小等于弦的张力，而其方向平行于弦的切线方向．

我们试考虑弦的某一小段．将弦上每点 x 离其平衡位置的距离记为 $u(x, t)$．我们假设弦的振动是在一个平面中进行的，且设这种位移垂直于 Ox 轴．我们把位移 $u(x, t)$ 用图形(图2)表示出来．我们考虑在点 x_1 和 x_2 之间的那一小段弦，于是有二力作用在这两点上，此力的大小等于弦的张力 T，而其方向平行于 $u(x, t)$ 的切线方向．

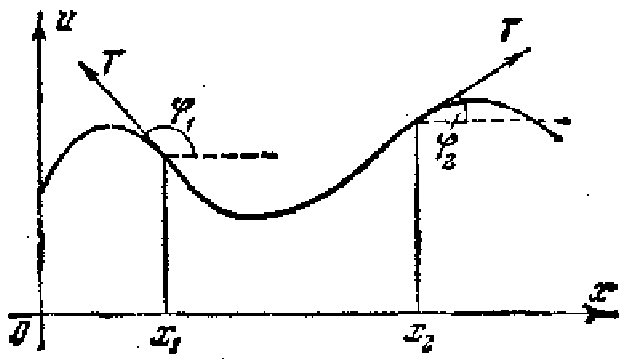


图 2

如果这弦是弯曲的，那么这二力的总作用不为零．根据力学的基本定律，这总作用力应该等于这段弦的动量的变化速度．

设每公分长的弦的质量为 ρ，则动量变化速度为

$$\rho\int_{x_1}^{x_2}\frac{d^2u}{dt^2}\,dx.$$

如果用 φ 表示弦的切线和 Ox 轴的夹角，我们便有

$$T\sin\varphi_2-T\sin\varphi_1=\int_{x_1}^{x_2}\partial\frac{d^2u}{dt^2}\,dx.$$

这就表示力学第二定律的基本方程的积分形式．不难把它表示为微分的形式．显然我们有

$$\rho\frac{d^2u}{dt^2}=\frac{\partial}{\partial x}(T\sin\varphi).$$

由微分学就容易看到未知函数 u 和 $T\sin\varphi$ 之间有下列的关系：

$$\operatorname{tg}\varphi=\frac{\partial u}{\partial x},\quad \sin\varphi=\frac{\operatorname{tg}\varphi}{\sqrt{1+\operatorname{tg}^2\varphi}}=\frac{\dfrac{\partial u}{\partial x}}{\sqrt{1+\left(\dfrac{\partial u}{\partial x}\right)^2}},$$

且近似地设 $\left(\dfrac{\partial u}{\partial x}\right)^2$ 很小，则有

$$\sin\varphi\approx\frac{\partial u}{\partial x}.$$

于是

$$T\frac{\partial^2u}{\partial x^2}=\rho\frac{\partial^2u}{\partial t^2}. \tag{6}$$

这个方程就是弦振动方程的微分形式．

数学物理方程的基本形式 我们在前面已经说过，描述物理现象的偏微分方程一般都是几个未知函数的方程组．但在许多情形下，这些方程组可以化为一个方程．我们举出下列几个简单的例子来说明这一点．

我们来看一下前节中所考虑过的运动方程，要解这些方程必须同时考虑连续性方程，关于如何做法以后要谈到．

1．我们首先考虑理想流体的定常流动的方程．

流体的所有可能的运动可以分为有旋运动和无旋运动，无旋运动也叫作场位运动．虽然无旋运动只是流体运动的一部分情形，而且在某种意义之下，液体或气体的运动总是有旋的，但是经验说明，在许多情形下，这些无旋运动十分精确地体现着流体的运动．此外，从理论上可以证明：在无粘性的流体中，如果一开始没有旋流，那么以后也就不会产生．

对于流体的场位运动，存在一个纯量函数 $U(x, y, z, t)$，叫作**速度势**，速度向量 $\boldsymbol{v}$ 可以表示为

$$v_x=\frac{\partial U}{\partial x},\ v_y=\frac{\partial U}{\partial y},\ v_z=\frac{\partial U}{\partial z}.$$

直到目前为止，在所有的情形下，我们遇到的是四个未知函数的四个方程的方程组，也可以说是一个纯量的方程和一个向量的方程，其中包含一个未知的纯量函数和一个未知的向量场．通常这些方程可以化为关于一个未知函数的一个方程，但为二阶的．我们先就简单的情况说起．

对于不可压缩液体的场位运动，因 $\frac{\partial \rho}{\partial t}=0$，故有下列的两组方程：连续性方程

$$\rho\left(\frac{\partial v_x}{\partial x}+\frac{\partial v_y}{\partial y}+\frac{\partial v_z}{\partial z}\right)=0$$

和场位运动方程

$$v_x=\frac{\partial U}{\partial x},\ v_y=\frac{\partial U}{\partial y},\ v_z=\frac{\partial U}{\partial z}.$$

将这第二组方程表示出来的速度的数值代入第一个方程便得

$$\frac{\partial^2 U}{\partial x^2}+\frac{\partial^2 U}{\partial y^2}+\frac{\partial^2 U}{\partial z^2}=0. \tag{7}$$

2. 也可以通过微商的关系，将"热流"的向量场和一纯量——温度联系起来．大家知道热量总是从物体较热的部分"流"向较冷的部分．所以可以认为热流向量与温度的梯度向量相反．很自然，第一步近似地假设这向量的数值与所谓温度的梯度成正比．实验证明这假设是正确的．

温度的梯度向量有分量

$$\frac{\partial T}{\partial x},\ \frac{\partial T}{\partial y},\ \frac{\partial T}{\partial z}.$$

命比例系数为 k，便得下列三方程

$$\tau_x=-k\frac{\partial T}{\partial x},\ \tau_y=-k\frac{\partial T}{\partial y},\ \tau_z=-k\frac{\partial T}{\partial z}.$$

将这些数值代入热能的平衡方程

$$C\frac{\partial T}{\partial t}+\frac{\partial \tau_x}{\partial x}+\frac{\partial \tau_y}{\partial y}+\frac{\partial \tau_z}{\partial z}=q,$$

则得

$$C\frac{\partial T}{\partial t}=k\left[\frac{\partial^2 T}{\partial x^2}+\frac{\partial^2 T}{\partial y^2}+\frac{\partial^2 T}{\partial z^2}\right]+q. \tag{8}$$

3. 最后，对于气体介质的微小振动，譬如声的振动，由方程

$$\frac{\partial^2\rho}{\partial t^2}+\rho\frac{\partial}{\partial x}\left(\frac{dv_x}{dt}\right)+\rho\frac{\partial}{\partial y}\left(\frac{dv_y}{dt}\right)+\rho\frac{\partial}{\partial z}\left(\frac{dv_z}{dt}\right)=0$$

和运动方程(5)

$$\rho\frac{dv_x}{dt}+\frac{\partial p}{\partial x}=F_x,\ \rho\frac{dv_y}{dt}+\frac{\partial p}{\partial y}=F_y,\ \rho\frac{dv_z}{dt}+\frac{\partial p}{\partial z}=F_z,$$

假设没有外力($F_x=F_y=F_z=0$)，我们得

$$\frac{\partial^2 p}{\partial t^2}=a^2\left(\frac{\partial^2 p}{\partial x^2}+\frac{\partial^2 p}{\partial y^2}+\frac{\partial^2 p}{\partial z^2}\right) \tag{9}$$

(为此，只须将加速度的表达式代入连续性方程，且借助波义尔-马略特定律 $p=a^2\rho$ 将其中的密度 ρ 消去).

不仅对以上所考虑的问题，而且对许多数学物理问题而言，方程(7)，(8)和(9)是典型的. 这些方程的详细研究使我们明确了许多物理现象.

§3. 始值条件和边值条件. 解的唯一性

和常微分方程一样，除去极少的例外，每个偏微分方程一般有无穷个特解. 所以在解决具体的物理问题时，即寻求适合某方程的一个未知函数时，必须会从无穷个互不相同的解中选取我们所需要的解. 为此，除方程以外，还须知道一些附加条件. 我们在上面已经看到，偏微分方程是物理或力学的规律的表示式. 仅仅知道这种规律还不足预言某种现象. 譬如在天文学中，要预言天体的运动，当然须知道这些天体的质量，并且除牛顿定律的一般公式以外，还须要知道我们所研究的天体系统的原始状态，即在某起初时间，这些天体的分布及其速度. 在解决任何数学物理的问题时，

总会有类似的附加条件.

所以，一个数学物理的问题乃是要寻求一适合某些附加条件的偏微分方程的解.

关于上面所写出来的方程(7)，(8)和(9)，由于它们互不相同的特点，于是对它们所可能提出的问题和解决的问题也是互不相同的.

拉普拉斯方程和波桑方程．调和函数及其边值问题的解的唯一性．现在我们比较详细地考虑上述问题的提法．首先考虑拉普拉斯方程和波桑方程．方程[1]

$$\Delta u = -4\pi\rho$$

叫做**波桑方程**，其中 ρ 一般有密度的意义．在特殊情况下，ρ 可能为零．当 $\rho=0$ 时，我们便得**拉普拉斯方程**

$$\Delta u = 0.$$

不难看到，任何波桑方程的两个特解之差是一个适合拉普拉斯方程的函数，或者说，它们的差是一个**调和函数**．这样一来，波桑方程的解的整个集合化为调和函数的集合.

如果我们会用任一方法来建立波桑方程一个特解 u_0，那么用公式

$$u = u_0 + w$$

来变换未知函数，则得到关于新的未知函数 w 的拉普拉斯方程．我们也可以将附加条件作相应的变换，而得到关于新未知函数的附加条件，所以关于拉普拉斯方程的问题的研究是特别重要的.

正如许多数学问题一样，偏微分方程问题的正确提法往往是直接从实际中提示出来的．解决拉普拉斯方程时所产生的附加条件也是由问题的物理提法导来的.

我们考虑某介质中定常的温度状态，即当热源保持不变时，考虑该介质中的温度分布情形．在这些条件之下，经过某一段时间

1) 为了简单起见，我们用 Δu 表示下列的表达式：

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2},$$

这表达式一般称为函数 u 的拉普拉斯算子.

以后，介质每一点的温度将不随时间而变化。欲知道每一点的温度，我们求方程

$$\frac{\partial T}{\partial t}=\Delta T+q$$

的解，其中 q 是热量输出的密度，此解与 t 无关。我们得到

$$\Delta T+q=0.$$

由此可见，我们的介质中的温度应适合波桑方程。如果热量输出的密度为零，则波桑方程就变成拉普拉斯方程。

由简单的物理概念可以看到：欲求出介质内部的温度，我们还必须知道介质边界上的情况。

显然，所有以前对介质内点考虑的物理规律，在介质的边界点上将有另外的表达形式。

在定常的温度分布问题中，在边界上，我们或者给出温度的分布，或者给出通过单位面积的热流量，或者给出温度和热流量之间的某个关系。

如果我们考虑体积 Ω 中的温度，设这体积的边界面是 S，那么这些条件可写成

1) $$T\big|_s=\varphi(Q),\tag{10}$$

或

2) $$\frac{\partial T}{\partial n}\bigg|_s=\psi(Q),\tag{10'}$$

或在更一般的情形下

3) $$\alpha\frac{\partial T}{\partial n}\bigg|_s+\beta T\bigg|_s=\chi(Q),\tag{10''}$$

其中 Q 表示曲面 S 上的任意点。形式(10)的条件叫做边值条件。在寻求拉普拉斯方程或波桑方程的解时，上述形式的一个边值条件就确定了我们所需要的解的选取，并且一般是唯一地确定了这个解。

所以当求拉普拉斯方程或波桑方程的解时，必须且只须给出边界上的一任意函数[1]。现在我们比较详细地考虑拉普拉斯方程。

1) "任意函数"是指在这函数上，除某种正则性条件以外，不加其他特殊的限制。

如果我们知道一个调和函数(即适合拉普拉斯方程的函数)在区域的边界上取什么数值,那末这调和函数就完全地被确定。

我们先证明，一个调和函数在区域内部任一点的数值不能大于它在边界上的数值的极大. 或更精确地说，任何调和函数在边界上达到它的绝对极大值和绝对极小值.

由此立刻看到，如果一调和函数在区域 Ω 的边界上的数值是常数，那末这调和函数在这区域内部也等于这同一个常数. 因为若一函数的极大和极小是同一个常数，则此函数到处都等于这个常数。

现在我们证明调和函数的绝对极大和绝对极小不能在区域的内部. 我们先注意：若函数 $u(x, y, z)$ 的拉普拉斯算子 Δu 在整个区域中是正的,则此函数在区域内部不能有极大;若 Δu 为负,则函数在区域内不能有极小. 诚然,设函数 u 在某一点取极大,则单独对每一个变数来说(将其余二变数保持不变)，函数在这点仍为极大. 由此可见，这函数对每个变数的二阶微商为非正值. 这就是说，这些二阶微商的和也是非正的，故拉普拉斯算子不可能是正的. 同样可以证明：若函数在某个内点上为极小,则其拉普拉斯算子在这一点不可能为负. 这就是说，若一函数的拉普拉斯算子在区域中到处为负,则此函数在这区域中不可能有极小.

若某函数是调和函数，则总可以将它作任意小的变化而使其拉普拉斯算子为正或为负;为此只须加上以下的项

$$\pm\eta r^2=\pm\eta(x^2+y^2+z^2),$$

其中 η 是任意小的常数.

一个函数是否在区域的内部有绝对的极大或绝对的极小，加以十分微小的量并不改变这种性质. 如果一调和函数在区域的内部有极大,则加以 $+\eta r^2$ 以后,我们便得一个具正拉普拉斯算子的函数,而这函数在区域的内部有极大，但根据以上的证明，这是不可能的. 这就是说,调和函数不可能在区域的内部达到绝对极大.

同样可以证明，调和函数也不可能在区域的内部达到它的绝对极小。

由以上所证明的定理可以推出一个重要结论：如果两个调和函数在区域的边界上取相同数值，则此二函数在整个区域的内部互相一致．诚然，这二函数之差（仍为调和函数）在区域的边界上为零，故在区域内到处为零．

我们看到，调和函数被它在边界上的数值所完全确定．可以证明（但不能详细地给出这个证明），这些值可以任意给，并且总可以找出一个调和函数取此边值．

如果知道通过物体表面的每个面积元素的热流量，或知道温度和热流量之间规律，则物体定常的温度分布就完全被确定了．这事实的证明比较复杂，待以后讨论数学物理问题的解法时，再回到这问题上来。

热传导方程的边值问题　在非定常的情形下，热传导方程的定解问题的提法与以上的是完全两样的．从物理上可以看到，仅仅知道边界的温度或仅知道边界上的热流量，并不足以确定问题的解．但若除此以外，还知道在某个开始时刻的温度分布，则问题就完全确定了．因此，欲寻求形如(8)的方程的解，一般必须且只须给出一任意函数 $T_0(x, y, z)$ 作为原始的温度分布，且在边界上还要给出一任意函数．和以前一样，这可以是物体表面上的温度分布，或是通过表面每个面积元素的热流量，或是温度和热流量之间的规律．

因此我们的问题的提法是：寻求方程(8)的解，它适合条件

$$T|_{t=0}=T_0(x, y, z) \tag{11}$$

以及下列三条件之一：

1) $$T|_s=\varphi(Q), \tag{12}$$

2) $$\left.\frac{\partial T}{\partial n}\right|_s=\psi(Q), \tag{12'}$$

3) $$\alpha\left.\frac{\partial T}{\partial n}\right|_s+\beta T|_s=\chi(Q), \tag{12''}$$

式中 Q 是曲面 S 上的任意点．

条件(11)叫始值条件，而条件(12)叫边值条件．

我们不详细证明上述每个问题都有确定的解，而只考虑其中的第一个问题，而且只考虑介质内部没有热量输出的情形。我们来证明，方程

$$\Delta T=\frac{1}{a^2}\frac{\partial T}{\partial t}$$

和条件

$$T|_{t=0}=T_0(x,\ y,\ z),$$

$$T|_s=\varphi(Q)$$

只可能有一个解。

这个命题的证明与上述拉普拉斯方程的解的唯一性证明十分相象。首先我们证明，若

$$\Delta T-\frac{1}{a^2}\frac{\partial T}{\partial t}<0,$$

则作为四个变元 $x,\ y,\ z$ 和 $t(0\leqslant t\leqslant t_0)$ 的函数 T 或者在变元 $x,\ y,\ z$ 的区域 Ω 的边界上达到极小(或在 Ω 内达到极小，但必须在起始时刻 $t=0$)。

诚然，在相反的情形，T 在某内点达到极小，于是在这一点上，T 的所有一阶微商，包括 $\frac{\partial T}{\partial t}$ 在内，将等于零。而若这极小出现在 $t=t_0$，则 $\frac{\partial T}{\partial t}$ 非正。在这一点，T 对变元 $x,\ y,\ z$ 的二阶微商皆非负。由此可见，$\Delta T-\frac{1}{a^2}\frac{\partial T}{\partial t}$ 将为非负。所以在我们的区域内部，不能达到极小值。

同样可以证明，若 $\Delta T-\frac{1}{a^2}\frac{\partial T}{\partial t}>0$，则在 Ω 内，于 $0<t\leqslant t_0$ 时，函数 T 不能有极大值存在。

最后若 $\Delta T-\frac{1}{a^2}\frac{\partial T}{\partial t}=0$，则在 Ω 内，当 $0<t\leqslant t_0$ 时，函数 T 不能达到绝对的极大或绝对的极小，因为，譬如 T 达到绝对极小，则加上一项 $\eta(t-t_0)$，而考虑 $T_1=T+\eta(t-t_0)$，当 η 十分小时，不失去绝对极小的存在性，但此时 $\Delta T_1-\frac{1}{a^2}\frac{\partial T_1}{\partial t}$ 为负，而这是不可能的。

用同样的方法可以证明，T 在所考虑的域内没有绝对极大.

所以，绝对极大和绝对极小只能在初始时刻 $t=0$ 达到，或在所考虑介质的边界 S 上达到．如在初始时刻和边界 S 上，$T=0$，则对一切 $t\leqslant t_0$，在区域内部到处有 $T=0$．如果任意两个温度分布 T_1 和 T_2 在初始时刻 $t=0$ 和在边界上相等，则其差 $T_1-T_2=T$ 将适合热传导方程，且于 $t=0$ 和边界上为零．这就是说，T_1-T_2 到处等于零，故这二温度分布 T_1 和 T_2 到处相同．

在以后讨论数理方程的解法时，我们可以看到，T 在 $t=0$ 上的数值，以及等式(12)右边函数，可以任意指定，即每个这种问题皆有解存在.

振动能和振动方程的边值问题　我们现在考虑，在什么条件下我们已导出的，基本的数学物理方程的第三个方程，即方程(9)是可解的.

为了简单起见，我们考虑弦振动方程 $\frac{\partial^2 u}{\partial x^2}=\frac{1}{a^2}\frac{\partial^2 u}{\partial t^2}$，它和方程(9)很相似，仅仅是在空间变元的个数上有所不同．这方程的右边是 $\frac{\partial^2 u}{\partial t^2}$，它表示弦上任一点的加速度．对任一力学系统来说，如果知道了这系统每一点的初始位置和速度，则此系统的运动将完全被确定．因此对弦振动方程来说，给出每点的初始位置和速度是自然的：

$$u\big|_{t=0}=u_0(x)$$

$$\left.\frac{\partial u}{\partial t}\right|_{t=0}=u_1(x).$$

此外，我们在上面已经说过，这方程只能表示弦的内点的力学规律，它在弦的端点上就不再成立了．所以在弦的两端必须给出补充条件．如果弦的两端固定在平衡位置上，则我们将有

$$u\big|_{x=0}=u\big|_{x=l}=0.$$

这些条件有时也换为另一些比较复杂的条件，但这种变化并不是本质的.

对方程(9)也可以提出一些与此类似的定解问题．欲定方程

(9)的一个解，一般给出

$$p|_{t=0}=\varphi_0(x,\ y,\ z), \tag{13}$$
$$\left.\frac{\partial p}{\partial t}\right|_{t=0}=\varphi_1(x,\ y,\ z),$$

以及下列"边值条件"之一：

$$p|_s=\varphi(Q), \tag{14}$$
$$\left.\frac{\partial p}{\partial n}\right|_s=\psi(Q), \tag{14'}$$
$$\alpha\left.\frac{\partial p}{\partial n}\right|_s+\beta p\Big|_s=\chi(Q). \tag{14''}$$

这与以前的不同之点在于：(11)中只有一个始值条件，而现在有两个始值条件(13)。

条件(14)显然表示在所考虑体积的边界上的物理规律。

一般来说，条件(13)与条件(14)中的任一个条件合起来就唯一地确定了问题的解。但我们在这里不证明这一事实。我们仅证明，对(14)中的一个条件，这样的解是唯一的。

如果我们知道某函数 u 适合方程

$$\frac{\partial^2 u}{\partial x^2}=\frac{1}{a^2}\frac{\partial^2 u}{\partial t^2}$$

和始值条件

$$u|_{t=0}=0,\ \left.\frac{\partial u}{\partial t}\right|_{t=0}=0$$

以及边值条件

$$\left.\frac{\partial u}{\partial n}\right|_s=0.$$

($u|_s=0$ 的情形也很容易研究.)

我们证明在这些条件之下，函数 u 将恒等于零。

为了证明这一性质，我们需要引进一个概念，利用这概念可以证明前面两个问题的解的唯一性。在这里可以用物理的思想来阐述这一概念。

我们所需要援引的一个物理定律是"能量守恒定律"为了简单起见，我们仍考虑振动着的弦，弦的每一点的位移 $u(x,t)$ 适合方程

$$T\frac{\partial^2 u}{\partial x^2}=\rho\frac{\partial^2 u}{\partial t_2}.$$

从 x 到 $x+\Delta x$ 的弦的每个振动部分的动能可表为

$$\frac{1}{2}\left(\frac{\partial u}{\partial t}\right)^2\rho\,dx.$$

如果弦不在平衡位置，则除动能以外，它还有位能．我们现在来计算这个位能．仍考虑从 x 到 $x+\Delta x$ 之间的这一小段弦．对 Ox 轴而言，这一小段的位置是倾斜，所以它的长度近似于

$$\sqrt{dx^2+\left(\frac{\partial u}{\partial x}dx\right)^2},$$

而其伸长量为

$$\sqrt{1+\left(\frac{\partial u}{\partial x}\right)^2}dx-dx\approx\frac{1}{2}\left(\frac{\partial u}{\partial x}\right)^2dx.$$

将这伸长量乘以张力 T，我们就得到这段被拉长了的弦的位能

$$\frac{1}{2}T\left(\frac{\partial u}{\partial x}\right)^2dx.$$

设弦长是 l，如果我们把弦的所有点的动能和位能加起来，就得到这弦的总能量

$$E=\frac{1}{2}\int_0^l\left[T\left(\frac{\partial u}{\partial x}\right)^2+\rho\left(\frac{\partial u}{\partial t}\right)^2\right]dx.$$

如果作用在弦的两端的力不作功，特别当弦的两端固定时，则弦的总能量应保持为一常数。

这样我们就得到方程

$$E=\text{const}.$$

这样得来的能量守恒定律的表达式是基本力学方程的一个数学推论，它可以从基本力学方程推导出来．因为我们已经把运动规律写成弦振动微分方程和端点上的条件，所以我们可以对能量守恒定律给一个数学证明．诚然，按普通运算规则，将 E 对时间求微商，便得

$$\frac{dE}{dt}=\int_0^l\left[T\frac{\partial u}{\partial x}\frac{\partial^2 u}{\partial x\partial t}+\rho\frac{\partial u}{\partial t}\frac{\partial^2 u}{\partial t^2}\right]dx.$$

利用波动方程(6)，将式中 $\rho\frac{\partial^2 u}{\partial t^2}$ 换为 $T\frac{\partial^2 u}{\partial x^2}$，则 $\frac{dE}{dt}$ 化为

$$\begin{aligned}\frac{dE}{dt}&=\int_0^l T\left[\left(\frac{\partial u}{\partial x}\frac{\partial^2 u}{\partial x\partial t}\right)+\frac{\partial u}{\partial t}\frac{\partial^2 u}{\partial x^2}\right]dx\\&=\int_0^l T\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\frac{\partial u}{\partial t}\right)dx\\&=T\frac{\partial u}{\partial x}\frac{\partial u}{\partial t}\Big|_{x=l}-T\frac{\partial u}{\partial x}\frac{\partial u}{\partial t}\Big|_{x=0}.\end{aligned}$$

若 $\frac{\partial u}{\partial x}\Big|_{x=0}$ 或 $u|_{x=0}$ 为零，且 $\frac{\partial u}{\partial x}\Big|_{x=l}$ 或 $u|_{x=l}$ 也为零，则

$$\frac{dE}{dt}=0,$$

这就证明了 E 是常数.

对波动方程(9)，也可以用同样的方法证明能量守恒定律。若 p 适合方程(9)和条件

$$p\Big|_s=0\quad 或\quad \frac{\partial p}{\partial n}\Big|_s=0,$$

则能量

$$E=\iiint\left[\left(\frac{\partial p}{\partial x}\right)^2+\left(\frac{\partial p}{\partial y}\right)^2+\left(\frac{\partial p}{\partial z}\right)^2+\frac{1}{a^2}\left(\frac{\partial p}{\partial t}\right)^2\right]dx\,dy\,dz$$

与时间 t 无关.

如果在初始的时刻，振动的总能量为零，那么它将恒为零，于是就不可能有运动发生。如果对于相同的始值条件和边值条件，波动方程有两个解 p_1 和 p_2，则其差 $v=p_1-p_2$ 将仍为波动方程的解，并且适合零边值条件和始值条件.

若我们计算函数 v 所描述的振动的"能量"，不难看到，这能量 $E(v)$ 在初始的时刻为零. 于是在任何时刻，此能量也仍为零，所以函数 v 恒等于零，这就是说原来的两个解 p_1 和 p_2 相同. 这就证明了问题的解的唯一性.

我们已经看到，对以上考虑的三个方程，我们所提出的问题都是正确的.

在上面的讨论中，我们也顺便研究了这些方程的解的一些简

点 $\xi=0$ 在每一时刻将对应于数值 $x=at$. 由此可见

$$\xi=x-at.$$

我们可以把任何函数 $u(x, t)$表示为

$$u(x, t)=\varphi(\xi, t),$$

对于解 u_1 我们有

$$u_1(x, t)=\varphi_1(\xi),$$

因此, 在这个坐标系统中, 解 $u_1(x, t)$是和时间没有关系的. 所以, 对以速度 a 运动着观察者来说, 这弦的变形是凝结不变动的. 而对于不移动的观察者来说, 则弦上有波沿着 Ox 轴的方向以速度 a 跑去.

同样, 解 $u_2(x, t)$可以看作是向相反的方向以速度 a 移动着的波.

对于无穷长的弦来说, 这两种波将无限地向远方传播. 当它们向不同的方向移动时, 它们可以互相叠置起来而形成奇特的图形. 两个波的相加的位移有时增大, 有时消失.

设 u_1 和 u_2 沿不同的方向向某一点移来, 如果它们的符号相同, 则位移增强; 如果它们的符号互异, 则位移减弱. 图 3 对两个特殊形式的波阐明了弦的一连贯的位置: 起初波相向而独立地

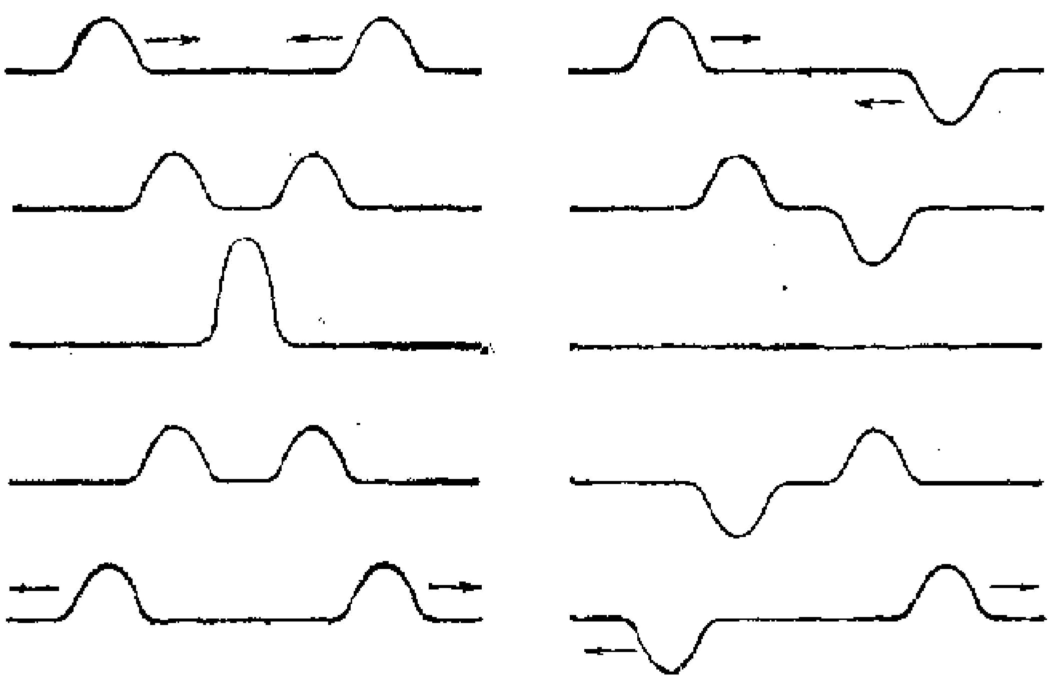


图 3

移动,继而开始互相接触．在第二种情况下,有一时刻振动完全消失,然后波又重新散开．所有的这些波是不难观察到的．

我们所要考虑的第二个范例是空间中的波的传播．

我们以前推导出来的方程

$$\Delta u=\frac{1}{a^2}\frac{\partial^2 u}{\partial t^2}, \tag{16}$$

具有下列形式的两个特解

$$u_1=\frac{1}{r}\varphi_1(r-at),\ u_2=\frac{1}{r}\varphi_2(t+at), \tag{17}$$

式中 r 表示从坐标原点到指定点的距离 $r^2=x^2+y^2+z^2$，而 φ_1 和 φ_2 表示任意二阶可微的函数．

欲验证 u_1 和 u_2 是解,需较多的时间,我们将不作这一验算．

这些解所描绘的波的图形一般和弦振动的情形一样．就(17)中的第一个解来说,如果不重视其因子 $\frac{1}{r}$，那末此解形成一个沿 r 增大的方向传播的波．这个波具有球对称性，即在离原点的距离相同的各点,这波的传播情形是完全一样的．

因子 $\frac{1}{r}$ 使波的振幅随其离原点的距离之增大成反比例地减小．这样的振动叫做发散的球面波．如果我们向水中投入一石子,则有圆形在水面上传播，这些传播着的圆可以看做是发散波的一种很好的表现;所不同的仅在于这是圆形波而不是球面波．

方程(17)中的第二个解也很有意思,它叫做收敛波,这波向着坐标原点的方向传播．其振幅逐渐增大,而到原点附近增至无穷．这种同心的球面波能使原来微小的振动聚集到一点而形成一个巨大冲击．

§5．解　　法

将任何解展成较简单的解的可能性　以上给出的数学物理问题的解可以用不同的方法得到．虽然这些问题的解法是完全特殊

的，但是它们有一个共同的思想基础. 我们看到，当未知函数值很小时，则所有数理方程对这些未知函数及其微商都是线性的，边值条件和始值条件也都是线性的.

同一个方程的任意两个解的差将仍为这一方程的解，唯方程的自由项为零. 这样的方程称为相应的纯方程.（例如，对波桑方程 $\Delta u=-4\pi\rho$ 来说，其相应的纯方程乃是拉普拉斯方程 $\Delta u=0$.）

如果同一个方程的两个解也适合相同边值条件，则其差适合相应的纯条件：即这差在边界上的相应的表达式的值为零.

因此，对指定的边值条件，方程的所有的解可以通过以下的方法得到：即把适合指定边值条件的任一特解加以纯方程而适合纯边值条件的一切可能的解（但一般不适合始值条件）.

将适合纯条件和纯方程的解相加或乘以一常数，仍得到纯方程的解，且适合纯条件.

如果适合纯条件和纯方程的解是某参数的函数，则对这参数积分，仍得适合纯条件和纯方程的解. 解决数理方程的所有线性问题的一种重要的方法——叠置法便是以这些事实为基础的.

我们试求形式为

$$u=u_0+\sum_k u_k$$

的解，式中 u_0 是适合边值条件而不适合始值条件的方程的一个特解，u_k 是相应的纯方程的一些解，它适合纯边值条件. 如果方程本来就是纯的，而且边值条件也是纯的，则可以求形式

$$u=\sum_k u_k$$

的解.

欲借纯方程的特解 u_k 的选取，使上述解能够适合任意始值条件，我们需要聚集足够多的这样的特解.

分离变数法 所谓分离变数法或傅里叶方法使我们能得到必要的上述特解.

通过范例

$$\Delta u=\frac{\partial^2 u}{\partial t^2}, \tag{18}$$

$$u|_s=0,\ u|_{t=0}=f_0(x,\ y,\ z),\ \left.\frac{\partial u}{\partial t}\right|_{t=0}=f_1(x,\ y,\ z)$$

来详细地研究这个方法.

为了求方程的特解，我们首先假设所要求的函数 u 适合边值条件 $u|_s=0$，且可表为两函数之乘积，其中一函数只与时间 t 有关，而另一个只与空间变量有关，

$$u(x,\ y,\ z,\ t)=U(x,\ y,\ z)T(t).$$

将这个解代入方程，便得

$$T(t)\Delta U=T''(t)U,$$

将这等式的两边除以 TU，乃得

$$\frac{T''}{T}=\frac{\Delta U}{U}.$$

这等式的右边是只与空间变量有关的函数，而左边与空间变量无关. 由此可见，欲使这等式成立，这等式的左右两边必为常数. 于是我们就得下列两个方程的方程组：

$$\frac{T''}{T}=-\lambda_k^2,\quad \frac{\Delta U}{U}=-\lambda_k^2.$$

为了表明所得到的常数为负(这可以严格证明)，我们用 $-\lambda_k^2$ 来表示等式右边的常数. 标数 k 用来表明 $-\lambda_k^2$ 的一切可能的数值可以有无穷多个，而且在某种意义之下，它们所对应的解构成一完备的函数族.

在上述两方程中除去分母，我们就得到

$$T''+\lambda_k^2T=0,\ \Delta U+\lambda_k^2U=0.$$

我们知道，式中的第一个方程有解

$$T=A_k\cos\lambda_kt+B_k\sin\lambda_kt;$$

这里 A_k 和 B_k 是任意常数. 可以引入辅助角 φ，而将这解作更进一步的简化，命

$$\frac{A_k}{\sqrt{A_k^2+B_k^2}}=\sin\varphi_k,\ \frac{B_k}{\sqrt{A_k^2+B_k^2}}=\cos\varphi_k,\ \sqrt{A_k^2+B_k^2}=M_k.$$

于是 $T=\sqrt{A_k^2+B_k^2}\sin(\lambda_kt+\varphi_k)=M_k\sin(\lambda_kt+\varphi_k)$. 函数 T 是以 λ_k 为频率以 φ_k 为相角的简谐振动.

对于纯边值条件，例如条件

$$U|_s=0$$

(式中 s 是考虑域 Ω 的边界)，
求方程

$$\Delta U+\lambda_k^2 U=0 \tag{19}$$

的解的问题是比较困难而有兴趣的问题．这问题的解并不总可以用已知函数的有限形式得到，但这些解恒存在，而且可以任意精确地求出．

于条件 $U|_s=0$ 下，方程 $\Delta U+\lambda_k^2 U=0$ 有显然解 $U\equiv 0$．这个解是不足道的，而且对我们毫无用处．若 λ_k 是随便取定的一常数，则我们的问题一般没有其他的解存在．但确实有这样的数值 λ_k 存在，对于这些数值，方程有非零解．

根据下列的要求可以确定一切可能的常数值 λ_k^2：即对每个这样的值，方程(19)有非零解，适合条件 $U|_s=0$．（由此也可推出，用 $-\lambda_k^2$ 表示的数值是负的．）

对于每个可能的数值 λ_k，由方程(19)至少可以找到一个函数 U_k．用它可以得到方程(18)的一个特解，

$$u_k=M_k\sin(\lambda_k t+\varphi_k)U_k(x,\ y,\ z).$$

这样的解称为我们所考虑的区域的**固有振动**．常数 λ_k 给出这固有振动的频率，而函数 $U(x,\ y,\ z)$给出这固有振动的形状．这函数叫做**固有函数**．在任何时刻，若把函数 u_k 看做是变量 $x,\ y,\ z$ 的函数，则它与 $U_k(x,\ y,\ z)$只差一个比例．

现在我们不可能详细证明固有振动和固有函数的所有的重要性质，我们仅仅将这些性质列举于下．

固有振动的第一个性质：对于任何体积，有无穷多个固有振动的频率，或称为**固有频率**．当标号 k 增大时，这些频率趋于无穷．

固有振动的第二个性质叫**直交性**．对应不同的 λ_k 的两固有函数的乘积在区域 Ω 上的积分为零[1)]

---

1) 如果有若干个本质上不相同的(线性无关的)函数 U 对应于同一数值 λ，则这个数值 λ 在集合 λ_k 中就出现相当多次．对应于相同的 λ_k 的函数的直交性条件可以由这些函数的适当选取来保证．

$$\iiint_{\Omega} U_k(x, y, z) U_j(x, y, z) \, dx\, dy\, dz = 0 \quad (j \neq k).$$

当 $j=k$ 时，我们可以设

$$\iiint_{\Omega} U_k(x, y, z)^2 \, dx\, dy\, dz = 1.$$

只要将函数 $U_k(x, y, z)$ 乘以适当的常数，便可以使这等式成立，因为函数 U 乘以常数后仍为方程(19)和条件 $U|_s=0$ 的解.

最后，固有振动的第三个性质：如果不遗漏任一数值 λ_k，则任意函数 $f(x, y, z)$，只要它适合边值条件 $f|_s=0$ 且其一、二阶微商无间断，于是它总可以用固有函数 $U_k(x, y, z)$ 任意精确地来表示．任一这样的函数 $f(x, y, z)$ 可以表示为收敛的级数

$$f(x, y, z) = \sum_{k=1}^{\infty} C_k U_k(x, y, z). \tag{20}$$

固有函数的第三个性质使我们在原则上能把任意函数 $f(x, y, z)$ 表示为我们的问题的固有函数的级数形式，而用第二个性质我们又可以决定这级数的所有系数．诚然，将等式(20)的两边乘以 $U_j(x, y, z)$，且在整个区域 Ω 上积分，便得

$$\iiint_{\Omega} f(x, y, z) U_j(x, y, z) \, dx\, dy\, dz$$

$$= \sum_{k=1}^{\infty} C_k \iiint U_k(x, y, z) U_j(x, y, z) \, dx\, dy\, dz.$$

由于直交性，等式右边和号中所有 $k \neq j$ 的项皆为零，而 C_j 的系数等于 1．因此我们得到

$$C_j = \iiint_{\Omega} f(x, y, z) U_j(x, y, z) \, dx\, dy\, dz.$$

以上所列举的固有振动的性质使我们可以解决振动的一般问题，问题的始值条件是任意的.

为此，我们考虑问题的解

$$u = \sum U_k(x, y, z) \, (A_k \cos \lambda_k t + B_k \sin \lambda_k t), \tag{21}$$

设法选取常数 A_k 和 B_k 使它适合

$$u|_{t=0}=f_0(x, y, z),$$
$$\left.\frac{\partial u}{\partial t}\right|_{t=0}=f_1(x, y, z).$$
将 $t=0$ 代入(21)的右边，我们看到，正弦的项皆为零，而 $\cos\lambda_k t$ 等于 1，所以我们有

$$f_0(x, y, z)=\sum_{k=1}^{\infty}A_k U_k(x, y, z).$$

根据固有振动的第三个性质，我们看到这种表示是可能的，又由第二个性质，我们有

$$A_k=\iiint_{\Omega}f_0(x, y, z)U_k(x, y, z)\,dx\,dy\,dz.$$

同样，将公式(21)对 t 微分，且代入 $t=0$，我们有

$$\begin{aligned}\left.\frac{\partial u}{\partial t}\right|_{t=0}&=f_1(x, y, z)\\&=\sum_{k=1}^{\infty}\lambda_k(B_k\cos\lambda_k t-A_k\sin\lambda_k t)\Big|_{t=0}U_k(x, y, z)\\&=\sum_{k=1}^{\infty}\lambda_k B_k U_k(x, y, z).\end{aligned}$$

和以上一样，由此可得

$$B_k=\frac{1}{\lambda_k}\iiint_{\Omega}f_1(x, y, z)U_k(x, y, z)\,dx\,dy\,dz.$$

知道了 A_k 和 B_k 我们就知道每个固有振动的振幅和相角.

这样我们就证明了，将固有振动叠加起来可以得到纯边值条件的问题的一般解.

因此所有解都可以用固有振动组成；并且若知道了始值条件，便可以求得这每个固有振动的振幅和相角.

在自变量较少的情形的振动也可以用同样的方法来研究．我们考虑两端固定的弦的振动作为一例，弦振动方程是

$$\frac{\partial^2 u}{\partial t^2}=a^2\frac{\partial^2 u}{\partial x^2}.$$

设弦长为 l，而其两端固定

$$u|_{x=0}=u|_{x=l}=0,$$

我们现在来求这问题的解. 仍然求特解

$$u_k=T_k(t)U_k(x).$$

和以前的推理一样,显然

$$T_k''U_k=a^2U_k''T_k,$$

或

$$\frac{T_k''}{T_k}=a^2\frac{U_k''}{U_k}=-\lambda_k^2.$$

由此

$$T_k=A_k\cos\lambda_k t+B_k\sin\lambda_k t,$$

$$U_k=M_k\cos\frac{\lambda_k}{a}x+N_k\sin\frac{\lambda_k}{a}x.$$

我们用边值条件来求 λ_k 的值,并不是对所有 λ_k 的值,端点的两个条件皆可以适合. 由条件 $U_k|_{x=0}=0$ 得 $M_k=0$, 即 $U_k=N_k\sin\frac{\lambda_k}{a}x$. 代入 $x=l$, 得 $\sin\frac{\lambda_k l}{a}=0$. 这必须 $\frac{\lambda_k l}{a}=k\pi$, 式中 k 是整数. 这就是说

$$\lambda_k=\frac{ak\pi}{l}.$$

由条件 $\int_0^l U_k^2\,dx=1$ 得 $N_k=\sqrt{\frac{2}{l}}$.

最后,我们有

$$U_k(x)=\sqrt{\frac{2}{l}}\sin\frac{k\pi x}{l},\ T_k=A_k\cos\frac{ak\pi t}{l}+B_k\sin\frac{ak\pi t}{l}.$$

因此我们看到, 弦的固有振动在整个弦上具有整数个正弦半波的形式. 每个固有振动皆有其一定频率, 并且这些频率可以按其增大的顺序排列为

$$\frac{a\pi}{l},\ 2\frac{a\pi}{l},\ 3\frac{a\pi}{l},\ \cdots,\ k\frac{a\pi}{l},\ \cdots.$$

大家知道, 在音弦振动时, 我们听到的就是这些频率. 频率 $\frac{a\pi}{l}$ 称为基音频率, 而所有其余的是所谓泛音频率. 固有函数 $\sqrt{\frac{2}{l}}\sin\frac{k\pi x}{l}$ 在区间 $0\leqslant x\leqslant l$ 中改变 $k-1$ 次符号. 诚然, $\frac{k\pi x}{l}$ 从 0 变到 $k\pi$, 故其正弦改变 $k-1$ 次符号. 固有函数等于零的点

称为振动的波节.

譬如,我们使弦在第一泛音的波节所对应的点处固定不动,则基音就消失了,我们只听到第一泛音,它就提高了八度音. 在演奏乐器(如小提琴、中提琴和大提琴)时,常使用这种方法.

在关于固有振动的问题的例子上,我们已讨论了分离变数法. 但是这方法具有极广泛的应用领域: 它可以用来解决热传导问题以及很多其他问题.

对热传导方程

$$\Delta T=\frac{\partial T}{\partial t}$$

和条件

$$T|_s=0,$$

和以前一样,我们将有

$$T=\sum F_k(t)U_k(x,\ y,\ z).$$

于是

$$\frac{F_k'(t)}{F_k(t)}=-\lambda_k^2,\ \Delta U_k+\lambda_k^2U_k=0.$$

所以解具有形式

$$T=\sum_{k=1}^{\infty}e^{-\lambda_k^2t}U_k(x,\ y,\ z).$$

这个方法可以用来解决其它的一些方程. 例如,我们在圆

$$x^2+y^2\leqslant 0$$

内考虑拉普拉斯方程

$$\Delta u=0,$$

如果我们所要寻求的解适合条件

$$u|_{r=1}=f(\theta),$$

式中 r 和 θ 表示平面上的点的极坐标.

拉普拉斯方程不难化为极坐标的形式

$$\frac{\partial^2u}{\partial r^2}+\frac{1}{r}\ \frac{\partial u}{\partial r}+\frac{1}{r^2}\ \frac{\partial^2u}{\partial\theta^2}=0.$$

我们求这方程呈下列形式的解:

$$u=\sum_{k=1}^{\infty}R_k(r)\theta_k(\theta).$$

欲级数的每一项皆单独地适合方程，我们就得到

$$\left[R_k''(r)+\frac{1}{r}R_k'(r)\right]\theta_k(\vartheta)+\frac{1}{r^2}\theta_k''(\vartheta)R_k(r)=0.$$

将方程除以 $\frac{R_k\theta_k(\vartheta)}{r^2}$，我们得到

$$\frac{r^2\left[R_k''(r)+\frac{1}{r}R_k'(r)\right]}{R_k(r)}=-\frac{\theta_k''(\vartheta)}{\theta_k(\vartheta)}.$$

再命

$$\frac{\theta_k''(\vartheta)}{\theta_k(\vartheta)}=-\lambda_k^2,$$

于是

$$r^2\left[R_k''+\frac{1}{r}R_k'\right]-\lambda_k^2R_k=0.$$

不难看到，函数 $\theta_k(\vartheta)$ 应为 ϑ 的周期函数，以 2π 为周期．求方程 $\theta_k''(\vartheta)+\lambda_k^2\theta_k(\vartheta)=0$ 的积分，我们得到

$$\theta_k=a_k\cos\lambda_k\vartheta+b_k\sin\lambda_k\vartheta.$$

欲这函数以 2π 为周期，λ_k 必须是整数，命 $\lambda_k=k$，我们便有

$$\theta_k=a_k\cos k\vartheta+b_k\sin k\vartheta.$$

R_k 的方程有形式为

$$R_k=Ar^k+\frac{B}{r^k}$$

的一般解．若只保留当 $r\to0$ 为有界的那些项，我们就得到了拉普拉斯方程的一般解

$$u=a_0+\sum_{k=1}^{\infty}(a_k\cos k\vartheta+b_k\sin k\vartheta)r^k.$$

这个方法也常常用来寻求方程 $\Delta U_k+\lambda_k^2U_k=0$ 适合纯边值条件的非零解，如果用这种方法可以把问题化为解常微分方程的问题，我们就说这问题的变量可以完全分离．苏联数学家斯捷巴诺夫曾证明，仅仅在一些特殊的情形下，可以用傅里叶方法将变数完全分离．很早以前，数学家就已经知道了分离变数法．欧拉、伯努利和达朗贝尔实质上就用过这一方法．但是，我们已经说过，在许多情形这方法是不能采用的；那我们就要应用另一些方法．现在

我们就来谈这些方法.

位势法 这个方法的实质和以前一样，仍然是以特解的叠加而求一般形式的解. 现在用来作为基本特解的是在空间某些点上变为无穷的那些解. 我们现就拉普拉斯方程和波桑方程的例子来阐明这种方法.

命 M_0 是我们的空间中的某一点. 用 $r(M, M_0)$ 表示从这点 M_0 到某另一动点 M 的距离. 将 M_0 固定，则

$$\frac{1}{r(M, M_0)}$$

是动点 M 的函数. 不难验证，除去 M_0 点本身，这函数是变点 M 在整个空间中的调和函数[1)]，而在 M_0 点，这函数及其微商变为无穷. 这种形式的函数的和

$$\sum_{i=1}^{N} A_i \frac{1}{r(M, M_i)},$$

仍然是点 M 的调和函数，其中 $M_1, M_2, \cdots, M_N$ 是空间中的任意点. 这函数在所有点 M_i 有奇异性. 将点 $M_1, M_2, \cdots, M_N$ 任意稠密地分布在某区域 Ω 中，同时将系数 A_i 减小，我们可以将这表达式取极限而得到一新的函数

$$U=\lim \sum_{i=1}^{N} \frac{A_i}{r(M, M_i)}=\iiint_{\Omega} \frac{A(M')}{r(M, M')}\, d\Omega,$$

其中 M' 跑遍整个区域 Ω. 这种形式的积分叫做**牛顿位势**. 可以证明，这样得到的函数 U 适合方程 $\Delta U=-4\pi A$. 但我们不在这里证明这个等式.

牛顿位势具有简单的物理意义. 为了说明这个意义，我们先研究函数 $\dfrac{A_i}{r(M, M_i)}$.

这函数对坐标的偏微商是

$$A_i \frac{x_i-x}{r^3}=X,\quad A_i \frac{y_i-y}{r^3}=Y,\quad A_i \frac{z_i-z}{r^3}=Z.$$

将质量 A_i 置于 M_i 点上，则它对任何物体有一个引力，此力的方

1) 即适合拉普拉斯方程.

向是向着 M_i 点，其大小与距离的平方成反比．我们现在来把这力分解成分量．如果作用在单位质量的质点上的力是$\frac{A_i}{r^2}$，则这力方向和坐标轴方向的夹角余弦为$\frac{x_i-x}{r}$，$\frac{y_i-y}{r}$，$\frac{z_i-z}{r}$．因此，引自中心 M_i 作用在点 M 的单位质量的力的分量等于 X, Y, Z，这就是函数 $\frac{A_i}{r}$ 对坐标的偏微商．如果在某些点 $M_1, M_2, \cdots, M_N$ 放置质量，则在 M 点的单位质量的质点将经受一个引力，此力等于从个别点 M_i 作用于此点 M 的力的总和，换言之，

$$X=\frac{\partial}{\partial x}\sum\frac{A_i}{r(M,\ M_i)},\ Y=\frac{\partial}{\partial y}\sum\frac{A_i}{r(M,\ M_i)},$$

$$Z=\frac{\partial}{\partial z}\sum\frac{A_i}{r(M,\ M_i)},$$

取极限，并将和号换成积分号，乃得

$$\overline{X}=\frac{\partial U}{\partial x},\ \overline{Y}=\frac{\partial U}{\partial y},\ \overline{Z}=\frac{\partial U}{\partial z},$$

式中 $U=\iiint\limits_{\Omega}\frac{A}{r}\,d\Omega.$

函数 U 的偏微商等于作用力的分量，这函数称为该力的位势．因此，函数 $\frac{A_i}{r(M,\ M_i)}$ 是 M_i 点的引力的位势，函数 $\sum\frac{A_i}{r(M,\ M_i)}$ 是诸点 $M_1, M_2, \cdots, M_N$ 的引力的位势，而函数 $U=\iiint\limits_{\Omega}\frac{A}{r}d\Omega$ 是连续分布在体积 Ω 中的质量的引力的位势．

质量不但可以分布在某个体积中，而且我们可以把这些点 $M_1, M_2, \cdots, M_N$ 分布在某个曲面 S 上．仍然把这些点的数目增大，并取极限，我们便得到积分

$$V=\iint\limits_{s}\frac{A(Q)}{r}\,ds, \tag{22}$$

式中 Q 是曲面 S 上的点．

不难看到，在曲面 S 的外部或曲面 S 的内部，这函数到处是

调和的．可以证明，这函数在曲面 S 上是连续的，但其一阶微商有间断．

对固定的 M_i，函数 $\dfrac{\partial\frac{1}{r}}{\partial x_i}$，$\dfrac{\partial\frac{1}{r}}{\partial y_i}$，$\dfrac{\partial\frac{1}{r}}{\partial z_i}$ 仍然是点 M 的调和函数．由这些函数，也可以作它们的和

$$\sum A_i\frac{\partial\frac{1}{r}}{\partial x_i}+\sum B_i\frac{\partial\frac{1}{r}}{\partial y_i}+\sum C_i\frac{\partial\frac{1}{r}}{\partial z_i},$$

除去点 $M_1, M_2, \cdots, M_N$ 以外，这函数到处是调和的．

我们仍然可以将点 $M_1, M_2, \cdots, M_N$ 的数目增加，作这些和的极限．积分

$$\begin{aligned}W&=\iint_s\mu(Q)\left(\frac{\partial\frac{1}{r}}{\partial x'}\cos(\boldsymbol{n},\ \boldsymbol{x})+\frac{\partial\frac{1}{r}}{\partial y'}\cos(\boldsymbol{n},\ \boldsymbol{y})\right.\\&\quad\left.+\frac{\partial\frac{1}{r}}{\partial z'}\cos(\boldsymbol{n},\ \boldsymbol{z})\right)ds\\&=\iint_s\mu(Q)K(Q,\ M)ds\end{aligned}\tag{23}$$

具有特殊的意义，在这积分中，x', y', z' 是曲面 S 上的变点 Q 的坐标；$\boldsymbol{n}$ 是曲面 S 在 Q 点的法线方向；$\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ 是坐标轴的方向；r 是从 Q 点到 M 点的距离，这样就确定了函数 W 在 M 点的数值．

积分(22)称为**单层位势**，积分(23)称为**双层位势**[1]．和单层位势一样，在曲面 S 的内部和外部，双层位势是调和函数．

调和函数论中的许多问题可以借助于位势来解决．关于在指

---

1) 位势的名称是和下列物理事实有关的．我们想象在曲面 S 上赋以电荷．于是这电荷便在空间引起一个电场．这个电场的位势将表示为积分(22)．所以这种积分(22)叫作单层位势．

如果曲面 S 是非导体的薄膜．在这薄膜的一側，我们按某种规律分布同符号的电荷(如正电荷)，而在另一側，仍按这规律分布反符号的电荷．这二电层的作用在空间形成一个电场．可以证明，这电场的位势可表示为积分(23)．

定区域 Ω 中求一调和函数 u，在这区域 Ω 的边界 S 上取指定边值 $2\pi\varphi(Q)$ 的问题，可以利用双层位势来解决。为了寻求这样的调和函数，只须选取适当的函数 $\mu(Q)$.

按其实质，这个问题好象是寻求级数

$$\varphi=\sum a_k U_k$$

的系数，使它成为右边部分。

积分 W 有一个非常好的性质，即当 M 点从曲面的里边趋近 Q_0 点时，这积分的数值为

$$\lim_{M\to Q_0} W=2\pi\mu(Q_0)+\iint_s K(Q,\ Q_0)\mu(Q)\,ds.$$

命这表达式等于已给的函数 $2\pi\varphi(Q_0)$，我们就得到方程

$$\mu(Q_0)+\frac{1}{2\pi}\iint_s K(Q,\ Q_0)\mu(Q)\,ds=\varphi(Q_0).$$

这方程称为**第二类积分方程**。关于这种方程，已经有许多很好的理论。随便用什么方法来解这个方程，那么就得到我们所考虑的问题的解。

用同样的方法可以得到调和函数论中其它问题的解。适当地择取位势，确定其密度，即确定其中的任意函数，使它适合所提出的条件.

从物理的观点来看，这就是说，如果在曲面 S 上以适当的密度分布这种双层电荷，那末所有调和函数都可以表为双层电位势.

近似解法．伽雷尔金方法和网络法 1．以前我们已经讲过了两种解决数理方程的方法：分离变数法和位势法．在十八世纪和十九世纪的学者(傅里叶、波桑、奥斯特洛格拉特斯基、李雅普诺夫等)的工作中，就已经研讨了这些方法．在二十世纪，这些方法又从其他许多方法得到了补充．我们现在来讲其中的两个方法——伽雷尔金方法和有限差分法．有限差分法也可以称为网络法．

第一种方法是伽雷尔金院士为了解决下列形式的方程而提出的：

$$\sum\sum\sum\sum A_{ijkl}\frac{\partial^4 U}{\partial x_i\partial x_j\partial x_k\partial x_l}+\sum\sum\sum B_{ijk}\frac{\partial^3 U}{\partial x_i\partial x_j\partial x_k}$$
$$+\sum\sum C_{ij}\frac{\partial^2 U}{\partial x_i\partial x_j}+\sum D_i\frac{\partial U}{\partial x_i}+EU+\lambda U=0,$$

这方程包含一个未知参数 λ(其中 i, j, k, l 互相独立地取 1, 2, 3 的值). 这个方程是从含自变量 t 的方程用分离变数法得来的, 正如从波动方程

$$\Delta u=\frac{\partial^2 u}{\partial t^2}$$

得到方程 $\Delta U+\lambda^2 U=0$ 一样. 我们的问题是要研究对什么样的数值 λ, 纯边值问题有非零解存在, 并且找到这些解.

伽辽尔金方法的实质可以简述如下: 该未知函数可以近似地表为

$$U\approx\sum_{m=1}^{N}a_m\omega_m(x_1, x_2, x_3),$$

式中 $\omega_m(x_1, x_2, x_3)$ 是一些适合边值条件的函数.

将它代入方程的左边, 就得到一个近似的等式.

$$\sum_{m=1}^{N}a_m\Big\{\sum\sum\sum\sum A_{ijkl}\frac{\partial^4\omega_m}{\partial x_i\partial x_j\partial x_k\partial x_l}$$
$$+\sum\sum\sum B_{ijk}\frac{\partial^3\omega_m}{\partial x_i\partial x_j\partial x_k}+\sum\sum C_{ij}\frac{\partial^2\omega_m}{\partial x_i\partial x_j}$$
$$+\sum D_i\frac{\partial\omega_m}{\partial x_i}+E\omega_m\Big\}+\lambda\sum_{m=1}^{N}a_m\omega_m\approx 0.$$

为了简单起见, 将花括弧中的表达式记作 $L\omega_m$, 于是这方程可写成

$$\sum a_m L\omega_m+\lambda\sum a_m\omega_m\approx 0.$$

在这近似等式的两边乘以 ω_n, 并且在我们所考虑的区域 Ω 上积分, 就得到

$$\iiint_\Omega\sum a_m\omega_n L\omega_m\,d\Omega+\lambda\iiint_\Omega\sum a_m\omega_m\omega_n\,d\Omega\approx 0$$

这等式还可以写成

$$\sum_{m=1}^{N}a_m\iiint_\Omega\omega_n L\omega_m\,d\Omega+\lambda\sum_{m=1}^{N}a_m\iiint_\Omega\omega_m\omega_n\,d\Omega\approx 0.$$

如果我们要这个等式准确地成立，于是我们就得到关于未知系数 a_m 的一次代数方程组．方程式的个数等于它的未知数的个数．欲使这方程组有异于零的解，则其行列式就必须为零．若将这行列式展开，我们就得到关于未知数 λ 的 N 次方程．

求得 λ 的数值，将它代入方程组，并且求这方程组的解，我们就得到了函数 U 的近似值．

伽雷尔金方法不仅可用于四阶方程，而且可以用于任何阶的各种类型的方程．

2．我们所要讲的最后一个方程是所谓有限差分法或网络法．

函数 u 对变量 x 的微商是比值

$$\frac{u(x+\Delta x)-u(x)}{\Delta x}$$

的极限．这个比值也可以表为

$$\frac{1}{\Delta x}\int_x^{x+\Delta x}\frac{\partial u}{\partial x_1}dx_1,$$

再根据大家所知道的中值定理(见第二章 § 8)

$$\frac{u(x+\Delta x)-u(x)}{\Delta x}=\frac{\partial u}{\partial x}\bigg|_{x=\xi},$$

式中 ξ 是区间

$$x<\xi<x+\Delta x$$

中的一点．

函数 u 的所有二阶微商，无论是混合微商，或是对一个变量的微商，也都可以近似地表示为差商的形式．诚然，差商

$$\frac{u(x+\Delta x)-2u(x)+u(x-\Delta x)}{(\Delta x)^2}$$

可以表为以下的形式

$$\frac{1}{\Delta x}\left\{\frac{u(x+\Delta x)-u(x)}{\Delta x}-\frac{u(x)-u(x-\Delta x)}{\Delta x}\right\}$$
$$=\frac{1}{\Delta x}\left\{\left[\frac{u(x_1+\Delta x)-u(x_1)}{\Delta x}\right]\Bigg|_{x_1=x-\Delta x}^{x_1=x}\right\}.$$

根据中值定理．函数

$$\varphi(x_1)=\frac{u(x_1+\Delta x)-u(x_1)}{\Delta x}$$

的差商可以用微商的数值来代替．所以

$$\frac{\varphi(x_1)-\varphi(x_1-\Delta x)}{\Delta x}=\varphi'(\xi),$$

式中 ξ 是区间

$$x-\Delta x<\xi<x$$

中的某一个中间值．

因此，

$$\left(\frac{1}{\Delta x}\right)^2[u(x+\Delta x)-2u(x)+u(x-\Delta x)]$$
$$=\frac{1}{\Delta x}[\varphi(x)-\varphi(x-\Delta x)]=\varphi'(\xi).$$

另一方面

$$\varphi(\xi)=\frac{u(\xi+\Delta x)-u(\xi)}{\Delta x},$$

这就是说

$$\varphi'(\xi)=\frac{u'(\xi+\Delta x)-u'(\xi)}{\Delta x}.$$

再应用一次上述的公式，便可以看到

$$\varphi'(\xi)=u''(\eta),$$

式中 $\xi<\eta<\xi+\Delta x$.

由此可见

$$\left(\frac{1}{\Delta x}\right)^2[u(x+\Delta x)-2u(x)+u(x-\Delta x)]=u''(\eta),$$

式中 $x-\Delta x<\eta<x+\Delta x$.

如果微商 $u''(x)$ 是连续的，且 Δx 的值充分小，则 $u''(\eta)$ 和 $u''(x)$ 的差将任意小，所以我们的二阶微商，可以任意近似于上面所考虑的差商．同样可以证明，混合微商

$$\frac{\partial^2 u}{\partial x\,\partial y}$$

可以近似地表示为公式

$$\frac{\partial^2 u}{\partial x\,\partial y}=\frac{1}{\Delta x\,\Delta y}[u(x+\Delta x,\ y+\Delta y)-u(x+\Delta x,\ y)$$
$$-u(x,\ y+\Delta y)+u(x,\ y)].$$

现在回到我们的偏微分方程上来.

为了确定起见，设我们考虑的方程是两个自变量的拉普拉斯方程

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0,$$

此外设未知函数 u 在区域 Ω 的边界 S 上的数值已经给定.

近似地取

$$\frac{\partial^2 u}{\partial x^2}=\frac{u(x+\Delta x,\ y)-2u(x,\ y)+u(x-\Delta x,\ y)}{(\Delta x)^2},$$

$$\frac{\partial^2 u}{\partial y^2}=\frac{u(x,\ y+\Delta y)-2u(x,\ y)+u(x,\ y-\Delta y)}{(\Delta y)^2}.$$

命 $\Delta x=\Delta y=h$, 于是

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=\frac{1}{h^2}[u(x+h,\ y)+u(x,\ y+h)$$

$$+u(x-h,\ y)+u(x,\ y-h)-4u(x,\ y)].$$

用以 $x=kh,\ y=bh$ 为顶点的正方形网络盖着区域 Ω(图 4).

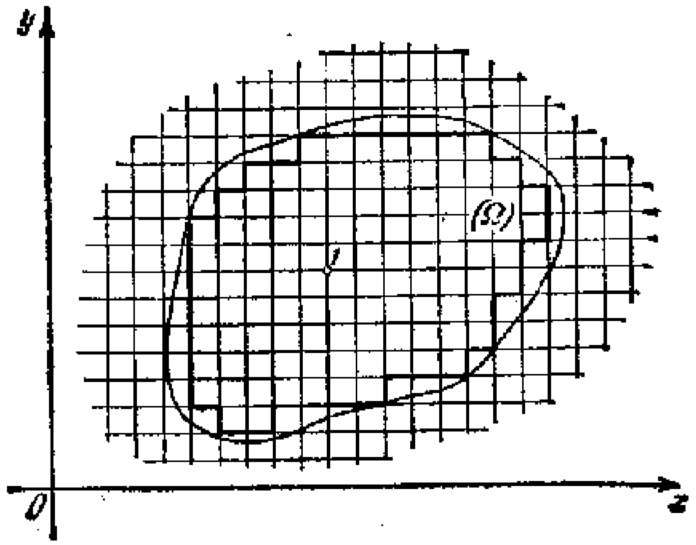


图 4

以网络在 Ω 中的正方形所构成的多角形来代替我们的区域 Ω, 以折线代替区域的边界. 将未知函数在边界 S 上的指定数值转移到这折线上来，且对所有内点用方程

$$u(x+h,\ y)+u(x,\ y+h)+u(x-h,\ y)$$
$$+u(x,\ y-h)-4u(x,\ y)=0$$

近似地代替拉普拉斯方程．这方程可改写为

$$u(x,\ y)=\frac{1}{4}[u(x+h,\ y)+u(x,\ y+h)$$
$$+u(x-h,\ y)+u(x,\ y-h)].$$

所以，函数 u 在网络的每一点上的数值，比如在图 4 中点 1 上的数值，等于它邻近四点上的数值的算术平均值．

如果在多角形的内部有 N 个网络的点．在每一个这样的点上，我们就得到一个方程．因此我们就得到了含有 N 个未知量的 N 个代数方程的方程组．解这方程组，我们便得到函数 u 在区域 Ω 内的近似值．

可以证明，对拉普拉斯方程来说，能以任意高的精确度得到它的解．

有限差分法将问题化为含 N 个未知量的 N 个方程的组来解决，而且把所求函数在某网络的结点上的数值取为未知量．

可以证明，这种差分方法可以用于数学物理的其他的问题，可用来解微分方程，也可用来解积分方程．但在若干情形，差分方法的应用遇到许多困难．

可以看到，按网络法得到的 N 个未知量的 N 个代数方程的组，或一般没有解存在；或即使存在，但与问题的真正的解相差很远．当解方程组时，如果误差越来越大，也就是说，如果把网络正方形的边长取得越小，方程就越多，若解决这方程组时得到的误差越大，那末就发生以上所述的情况．

在上面拉普拉斯方程的例中，不发生这样的情况，在按近似方法解这方程时，误差不但不越来越大，而相反地逐渐减小．对于热导方程或对于波动方程，基本网络的选取则非常重要．对这些方程既可以得到好的结果，也可能得到坏的结果．

如果我们要用网络法来解决其中的某一个方程，那末在选定 t 值的网络以后，则对空间变量就不能选取过分小的网络．否则

我们就要得到关于未知函数的数值的非常不好的方程组．解这方程组，将得到振动得很快的结果，而且振幅也很大．这个结果与真正的解相差很大．

从简单的数值计算的例子，可以看到各种不同的可能性．我们考虑方程

$$\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2},$$

在温度与 y 和 z 无关的时候，热导方程便化为这个方程．设网络沿 t 值的间隔是 k，而沿 x 值的间隔为 h，则

$$\frac{\partial u}{\partial t}\approx\frac{u(t+k,\ x)-u(t,\ x)}{k},$$

$$\frac{\partial^2 u}{\partial t^2}\approx\frac{u(t,\ x+h)-2u(t,\ x)+u(t,\ x-h)}{h^2}.$$

于是我们的方程可近似地写成形式

$$u(t+k,\ x)=\frac{k}{h^2}u(t,\ x+h)+\left(1-2\frac{k}{h^2}\right)u(t,\ x)+\frac{k}{h^2}u(t,\ x-h).$$

若知道 u 在某节时刻 t 的点 $x-h$, x, $x+h$ 的数值，我们就不难得到 u 在下一节时刻 $t+k$ 的点 x 的数值．命 k 是常数，既网络沿 t 的时间距已选定．我们考虑 h 的两种选取的情况．在第一种情况，命 $h^2=k$；在第二种情况，命 $h^2=2k$．用网络法来解决下列的问题．

在初始时刻，对 x 的所有负值，$u=0$；而对 x 的所有非负值，$u=1$．于是若将未知函数在固定时刻的数值写在同一横行上，我们就得到下面两个表．

表 1

t \ x	$-5h$	$-4h$	$-3h$	$-2h$	$-h$	0	h	$2h$	$3h$	$4h$	$5h$
0	0	0	0	0	0	1	1	1	1	1	1
k	0	0	0	0	1	0	1	1	1	1	1
$2k$	0	0	0	1	-1	2	0	1	1	1	1
$3k$	0	0	1	-2	4	-3	3	0	1	1	1
$4k$	0	1	-3	7	-9	10	-6	4	0	1	1
$5k$	1	-4	11	-19	26	-25	20	-10	5	0	1

表 2

$t \backslash x$	$-5h$	$-4h$	$-3h$	$-2h$	$-h$	0	h	$2h$	$3h$	$4h$	$5h$
0	0	0	0	0	0	1	1	1	1	1	1
k	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1	1	1	1	1
$2k$	0	0	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{4}$	1	1	1	1
$3k$	0	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{7}{8}$	$\frac{7}{8}$	1	1	1
$4k$	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{11}{16}$	$\frac{11}{16}$	$\frac{15}{16}$	$\frac{15}{16}$	1	1
$5k$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{13}{16}$	$\frac{13}{16}$	$\frac{31}{32}$	$\frac{31}{32}$	1

在表 2 中，对任意时刻，我们得到从一点到另一点均匀下降的数值．这个表给出了热导方程的解的很好的近似．从表面上看来，如果把 x 值的间隔分得越细，则精确度似乎应该越高．但相反，在表 1 中，u 的数值从正值到负值振动得很快，而且达到非常大的数值，超过始值很多．显然，在这个表中得到数值与对应于真正解的数值相差很远．

从这个例子可以看到，如果我们想用网络法得到真正解的充分近似的数值，而得到我们所希望的精确的结果，我们必须很慎重地选择网络的格式，并且通过事先的研究来证明这种方法的应用是正确的．

借助数理方程而得到的某自然科学问题的解，给出了该方程所描述的物理现象的形式和过程的数学描述．

因为在用数理方程建立现象的模型时，总不免要从这现象的许多方面加以抽象，抛弃许多次要的成分，而抽出那些似乎是主要的东西，所以这样得到的结果就不是绝对真实的．这些结果仅仅对我们所考虑的那些图象或模型是绝对正确的．这些图象总必须和实验相对照，而验核我们所考虑的现象模型是否近似于现象本身和充分精确地表示了现象本身．

因此，最后判定结果的真实性的只是实际的试验．但是仅仅在有了深刻的理论以后，才能正确提出和了解这些实际的试验．

在考察乐器上振动的弦时，只有了解了固有振动的叠加规律，我们才能理解乐器所发出的所有音调的起源．研究了频率如何被弦的质料、弦的张力以及其两端固定的特性所决定，我们才能理解频率之间的关系．在这情况，理论不仅给出某种数量的计算方法，并且提示出什么数量是具有特征的、物理过程是怎样进行的，并且在这物理过程中应该观察什么．

所以，从实际需要发生的科学领域数学物理，本身影响了实际并对实际提出了进一步发展的道路．

数学物理和数学分析的其它领域有密切的关系，但我们不能在这里介绍这些关系，否则我们将牵连得太远了．

§6. 广　义　解

在过去所考虑的问题中，现象都是以适合微分方程的连续可微函数来描写的，如果引入这些方程的间断解，那么问题范围可以有本质的推广．

在许多情形下，事先就知道我们所考虑的问题不可能有二阶连续可微的解，这就是说，从前几节所叙述的问题的古典的提法来看，这样的问题是没有解的．虽然如此，尽管在二阶可微函数类中不能找到描述这物理现象的函数，但仍有相当的物理现象发生．我们列举一些简单的例子于下：

1）如果弦是由两段不同密度的弦所构成的，则在方程

$$\frac{\partial^2 u}{\partial t^2}=a^2\frac{\partial^2 u}{\partial x^2} \tag{24}$$

中，系数 a 在相应的部分等于不同的常数，于是方程(24)一般就不可能有古典(二阶连续可微的)解．

2）设系数 a 为常数，但在开始的情形，弦子呈折线的形式，命此折线为 $u|_{t=0}=\varphi(x)$．在折线的顶点，函数 $\varphi(x)$ 一般没有一阶微商．可以证明方程(24)对于始值条件

$$u|_{t=0}=\varphi(x),\ u_t|_{t=0}=0$$

不可能有古典解$\left(今后以 u_t 表示 \frac{\partial u}{\partial t}\right)$．

3）如果打击弦的某一小部分，那么由这种作用所引起的振动将以方程

$$\frac{\partial^2 u}{\partial t^2}=a^2\frac{\partial^2 u}{\partial x^2}+f(x,\ t)$$

来描写．式中 $f(x,t)$ 相当于打击所引起的作用，它是一个间断函数，仅仅在一个短促的时间内在一小段弦上异于零．不难看到，这个方程也没有古典解．

由这些例子就已经看到，对未知解具连续微商的要求大大地缩小了可解问题的范围．为了寻找可解问题的较大的范围，首先允许被看做是问题的解的那些函数的最高阶微商可以有第一类的间断，而且除去间断线以外，这些函数到处适合方程．可以证明方程 $\Delta u=0$ 和 $\frac{\partial u}{\partial t}-\Delta u=0$ 的解在其定义域内不可能有这样的(所谓弱的)间断．而波动方程的解只可能在变量 x, y, z, t 的空间中的特殊形式的曲面(所谓特征曲面)上具有弱间断．如果于 $t=t_1$，把波动方程的解 $u(x, y, z, t)$ 看成是空间 x, y, z 中在时刻 $t=t_1$ 的一个纯量场．于是 $u(x, y, z, t)$ 的二阶微商的间断曲面将在空间 x, y, z 中移动，其速度等于波动方程中的拉普拉斯算子前的系数的平方根．

但是上述关于弦振动方程的第二个例说明，我们也必须考虑一阶微商有间断的解，而且例如在音波和光波的情况中，还必须考虑其本身就具有间断的那种解．在引起间断解时，我们要研究的第一个问题是：说明什么样的间断函数应该看作是各种方程以及关于这种方程的各种定解问题的解，而这种解是物理上可以接受的．譬如，我们是否可以认为任何分区为常数的函数是拉普拉斯方程或波动方程的“唯一的解”呢？因为它们除间断线以外都适合方程．

在明确这个问题的时候，必须首先考虑到：在包含所有允许解的那个最广的函数类中，唯一性定理应该仍然成立．很清楚，譬如我允许任意分区光滑的函数，那么这要求将不满足．

在起初，确定允许解的原则是要使这种函数为这方程的古典解(在某种意义下)的极限．比如，在以上引入的第二个例中，对应于函数 $\varphi(x)$ 的方程(24)的解在折点处没有微商，但是它可以看作是这方程(24)的古典解 $u_n(x,t)$ 的一致收敛的极限，$u_n(x,t)$ 是对应于始值条件 $u_n|_{t=0}=\varphi_n(x)$，$u_{nt}|_{t=0}=0$ 的解，式中 $\varphi_n(x)$ 是二阶连续可微的函数，当 $n\to\infty$ 时，一致收敛于 $\varphi(x)$．

在以后，又提出另一个原则来代替这个原则：即允许解 u 应该适合一含任意函数 Φ 的积分恒等式，来代替原来的方程 $Lu=f$．

这个积分恒等式是这样得来的：命 Φ 是和方程同阶连续可微的任意函数，它在方程定义域内的一有界域 D 的外部为零，将方程 $Lu=f$ 的两边乘以 Φ，将这样得到的等式在 D 上积分，然后利用分部积分使其中不含 u 的微商．结果得到我们所要的恒等式．对方程(24)来说，这积分恒等式为

$$\iint_{B} u\left(\frac{\partial^2 \Phi}{\partial t^2}-\frac{\partial^2 (a^2 \Phi)}{\partial x^2}\right) dx\, dt=0.$$

一个函数适合这样的积分恒等式，则称为广义解．索伯列夫对常系数方程证明了这两种原则界定义的允许解是等价的．但是对于变系数方程，第一个原则可能是不适用的．这些方程可能一般就没有古典解(参看例1)．在对方程系数的可微性质的极广的假设之下，第二个原则使我们可能确定广义解．诚然，初看起来，这个原则有些过于形式的纯数学的特点，而没有直接提示如何提出类似于古典问题的问题．

我们现在给出这原则的另一种叙述方式，它与大家所熟知的汉弥登原理有关，从而可以看到，这原则就物理学的观点来说是正确的．

大家知道，在十九世纪的前半叶，各种数理方程的分析推论引导出新规律——所谓汉弥登原理的发现，由这个原理出发，使我们可以用一致的方法得到所有已知的数理方程．我们用第三节中所考虑的关于固定端点的有界弦的振动问题例子，来阐述这个原理．

首先，对我们的弦建立所谓拉格朗日函数 $L(t)$，即动能和位能之差，由第三节可知

$$L(t)=\int_0^l\left[\frac{1}{2}\rho u_t^2-\frac{T}{2}u_x^2\right]dx.$$

设函数 $u(x, t)$ 对应于弦的真正的运动，$v(x, t)$ 在 $x=0$ 和 $x=l$ 上等于零，而在 $t=t_1$ 和 $t=t_2$ 时和 $u(x, t_1)$ 及 $u(x, t_2)$ 重合的其它函数，与这样的函数 $v(x, t)$ 相比较，根据汉弥登原理，对于函数 $u(x, t)$，积分

$$S=\int_{t_1}^{t_2} L(t)\, dt$$

取极小值，这里 t_1 和 t_2 是任意固定的两个数，而函数 v 应该有有限的积分 S．根据这个原则，S 的所谓第一变分(参看第八章)应该等于零，即

$$\delta S=\int_{t_1}^{t_2}\int_0^l [\rho u_t \Phi_t - T u_x \Phi_x]\, dx\, dt=0, \tag{25}$$

其中 $\Phi(x, t)$ 是任意对 x 和 t 的可微函数，而在长方形 $0\leqslant x\leqslant l$, $t_1\leqslant t\leqslant t_2$ 的边上等于零．

等式(25)便是那应该加于未知函数 $u(x, t)$ 的条件．如果知道 $u(x, t)$ 有二阶微商，则条件(25)可以有另一种形式．对积分(25)施行分部积分，且应用变分法基本引理，可知 $u(x, t)$ 应适合方程

$$\frac{\partial}{\partial t}\left(\rho\frac{\partial u}{\partial t}\right)-\frac{\partial}{\partial x}\left(T\frac{\partial u}{\partial x}\right)=0, \tag{26}$$

若 ρ 和 T 为常数，命 $a^2=\dfrac{T}{\rho}$，则这方程和(24)相同．

不难看到，对所有上述 Φ，方程(26)的任何解 $u(x, t)$ 适合恒等式(25)．反

之则不然，因为 $u(x, t)$一般可以没有二阶微商．因此，如果我们用恒等(25)来代替方程(26)，确乎把可解问题的范围扩大了．

为了确定弦振动的某个确定的规则，除边值条件

$$u(0, t)=u(l, t)=0 \tag{27}$$

以外，还必须提出始值条件

$$\begin{aligned} &u(x, 0)=\varphi_0(x), \\ &u_t(x, 0)=\varphi_1(x). \end{aligned} \tag{28}$$

如果是在一阶连续可微函数类中求解，则可以除(25)以外，独立地提出条件(27)和(28)，而把它们看成是补充条件．如果我们所要求的解"更不一些"，那么这些条件在上述的形式下就失去意义，于是应该把它们部分地或全部地包含在积分恒等式中．

譬如，命 $u(x, t)$在 $0\leqslant x\leqslant l,\ 0\leqslant t\leqslant T$ 中连续，而其一阶微商有间断．于是(28)中的第二个等式作为极限的条件，就失去了意义，在这种情况下，问题应该这样来提：即求一连续函数 $u(x,t)$，它适合条件(27)和条件(28)中的第一个条件，并且对所有在 $x=0, x=l, t=T$ 等于零的连续可微函数 Φ，等式

$$\int_0^T\int_0^l \left[\rho u_t\Phi_t - Tu_x\Phi_x\right] dx\,dt+\int_0^l \varphi_1\Phi(x, 0)dt=0 \tag{29}$$

恒成立，函数 u 和 Φ 应有一阶微商，且其平方在长方形 $0\leqslant x\leqslant l,\ 0\leqslant t\leqslant T$ 中勒贝格可积，对于 u 的这个限制的意思是说：弦的总能量按时间的中值

$$\frac{1}{2T}\int_0^T\int_0^l \left[\rho u_t^2+Tu_x^2\right]dx\,dt$$

是有限的．对函数 u 这种限制，以及在其可能的变化 Φ 上的限制，是汉弥登原则的自然推论．

恒等式(29)就是泛函

$$\tilde{S}=\int_0^T\int_0^l\left[\frac{\rho}{2}u_t^2-\frac{T}{2}u_x^2\right]dx\,dt+\int_0^l\varphi_1 u\Big|_{t=0}dx$$

的第一变分等于零的条件．因此在我们考虑的情况下，定弦的振动问题可以这样来提：即在所有适合条件(27)且于 $t=T$ 等于 $u(x, T)$的连续函数$v(x, t)$中，寻求泛函 $\tilde{S}$ 的极小．此外，所寻求的函数还需要适合(28)中第一个条件．

在这里所援引的汉弥登原理，不仅使我们可以扩大方程(24)的允许解的函数类，而使我们能对它们提出边值问题．

以上所引入的广义解或它的某些微商并不是对空间的所有点都可以确定的，但这并不与实验相矛盾．匈特曾以他的研究大大地促进了数理方程的解的新观点的形成，他不止一次地指出这一事实．

譬如，我们要确定水道中液体的流动，那么在古典的意义之下，就得计算

流体中每一点的流速向量和压力. 但在实际上，我们谈到的并不是流体在一点的压力，而总是它在某面积上的压力；也不说在指定点的速度向量，而说在单位时间内通过某面积流量. 广义解的定义，实质上提出了有直接物理意义那些量的计算.

欲使更多的问题可解，就得在尽可能广泛的函数类中找解；但是在这种广泛的函数类中，唯一性定理仍需成立. 这样的函数类往往接受了问题的物理实质. 譬如在量子力学中，作为许娄丁格方程的解的状态函数 $\psi(x)$是没有现实意义的；而有现实意义却是积分 $a_\nu=\int_E \psi(x)\psi_\nu(x)dx$，其中 ψ_ν 是某些平方可积的函数，即 $\int_E \psi_\nu^2 dx<\infty$. 因此不应该在二阶连续可微的函数类找解 $\psi(x)$，而应该在平方可积的函数类中求解. 在量子电动力学中，关于在什么样的函数类中寻求所考虑的方程的解的问题，到目前尚未彻底解决.

近三十年来数学物理的进展是和问题的这些新提法以及解决这些问题所必要的数学工具的创造有很多联系. 索伯列夫的所谓嵌入定理在这工具中占有中心的地位.

在这种或那种函数类中求广义解的特别方便的方法是有限差分法，变分直接法(里茨和脱莱夫茨方法)，伽雷尔金方法和泛函算子法由各种问题所引起的变换的性质的研究是这些方法的基础. 在§5中已谈到伽雷尔金方法和有限差分法. 在这里我们阐述变分直接方法的基本思想.

让我们考虑确定固定周界的弹性膜的平衡位置的问题，根据稳定平衡状态中位能的极小原理，设 $u(x, y)$是确定薄膜位置的函数，则在和 u 适合相同固定条件 $v|_s=\varphi$ 的所有其它连续可微函数 $v(x, y)$中，$u(x, y)$应该给出积分

$$J(u)=\iint_D (u_x^2+u_y^2)dx\,dy$$

的极小值. 如果边界 s 和 φ 满足某些限制. 可以证明这种极小是存在的，而且就是调和函数具备这样的性质，所以我们要寻找的函数 u 是狄立克雷问题 $\Delta u=0,\ u|_s=\varphi$ 的解. 反之亦然：和所有适合固定条件的函数 $v(x, t)$相比较，狄立克雷问题的解使得积分取极小值.

我们可以用里茨方法证明实现 J 的极小的函数 u 的存在性，并且能以任意精确的程度来计算这个函数 u. 我们取二阶连续可微函数的集合 $\{v_n(x, y)\}$，$n=0, 1, 2, \cdots$，当 $n>0$ 这些函数在边界上为零，而 $v_0(x, y)$在边界上等于 φ. 我们对呈形式

$$v=\sum_{k=1}^{n} C_k v_k+v_0$$

的函数，考虑积分 J，其中 n 是固定的，而 C_k 是任意数，于是 $J(v)$将为 n 个

未知变量 $C_1, C_2, \cdots, C_n$ 的二次多项式，我们确定 C_k 使这多项式取极小值，于是就得到含有 n 个未知量的 n 个线性代数方程的方程组，而其行列式异于零．所以 C_k 可以唯一确定．我们用 $v^n(x, y)$ 表示其相应的 v，如果 $\{v_n\}$ 适合某个"完备性"条件，则当 $n \to \infty$ 时，v^n 将收敛于某个函数，这函数就是我们所要找的问题的解．

最后，我们要指出，在这一章中仅仅对最简单的线性的力学问题作了一个描述，还有许多与更一般的偏微分方程有关的、远没有研究彻底的问题还没有提到，而这些问题还远没有得到彻底的研究．

文 献

Р. Курант и Д. Гильберт, Методы математической физики, Т. I—II. Перев. с нем. Гостехиздат, 1951 (I 有中译本).

在这本书中包含很多数学物理的方法．

В. И. Смирнов, Курс высшей математики, Т. II. Изд. 14—е. Гостехиздат, 1956. (有中译本).

这一卷中有一章叙述了最简单的数学物理问题的解法．

Ф. Франк и Р. Мизес, Дифференциальные и интегральные уравнения математической физики, Перев. с нем. ОНТИ, 1937.

这本书是由若干作者根据原来黎曼讲座的讲义而写成的，其中给出了物理各部分许多问题的数学提法和与这些问题有关的方程的研究．

近代大学课程(均有中译本)

И. Г. Петровский, Лекции об уравнениях с частными производными, Изд. 2—е, Гостехиздат. 1953.

В. И. Смирнов, Курс высшей математики, Т. IV, Изд. 3—е, Гостехиздат, 1953.

С. Л. Соболев, Уравнения математической физики, Изд. 3—е. Гостехиздат, 1954.

专 门 著 作

О. А. Ладыженская, Смешанная задача для гиперболического уравнения, Гостехиздат, 1953.

С. Г. Михлин, Проблема минимума квадратичного функционала, Гостехиздат, 1952.

С. Л. Соболев, Некоторые применения функционального анализа в математической физике, Изд. ЛГУ. 1950.

王光寅 译

周毓麟 校

第七章 曲线和曲面

§1. 关于曲线和曲面理论的对象和方法的概念

在中学几何课程里只研究最简单的线：直线、折线、圆和圆弧；而对于面，除掉讨论平面以外，只讨论多面体、球、圆锥、圆柱等的表面．在范围更广的课程中还研究另一些曲线，首先是圆锥截线：椭圆、抛物线、双曲线．但是研究任意的曲线和曲面总完全不是初等几何的事实．初看起来好像不能说出，对任意的曲线和曲面，可以举出和探讨些什么样的普遍性质．然而这种研究却是十分自然和必要的．

在实践活动和自然知识的所有方面，我们经常遇到各种形状的曲线和曲面．行星在空间中的路线，轮船在海上的路线，炮弹在空中的路线，切削刀具在被加工的金属上留下的痕迹，车轮在公路上留下的痕迹，笔尖在自动记录器纸带上的痕迹，操纵发动机活塞的凸轮的周界，美术图案的花纹，悬挂着的缆索的形状，特殊绕法的螺旋弹簧的形状，诸如此类的各种曲线是说不完的．立体的表面、薄壳、油罐、飞机的外壳、套子、金属薄片等等，给出各种各样的曲面．物品加工的方法、光学性质、立体的流线型、薄壳的刚性、坚固性和可变性以及许多其他的性质，在很大程度上都依赖于物体表面的几何形状．

当然，切削刀具在被加工的金属上留下的小道并非数学上的曲线．即使是四壁极薄的油罐，也不是数学上的曲面．但是把现实的对象设想成数学的曲线和曲面，即使只是最初的近似，已经足以用于研究许多问题了．

现实的线总是有粗细的，我们是在舍弃了限制粗细程度的所

有因素的基础上引起数学曲线的概念的，因此我们把曲线设想成绝对地细的线、没有粗细的线．我们从这个抽象概念所反映出来的是物体的完全现实的普遍性质，这种性质是物体在其厚度和宽度比其长度小得多的情况下所具有的．

同样地，舍弃了在减少薄壳厚度时所受到的限制和在确定物体表面的精确位置时所受到的限制，我们就引出了数学曲面的概念．我们不预备给出这些极其熟知的概念的严格定义．我们只是指出一下，它们的精确的数学定义并不简单，而且是属于拓扑学的范围的．

最后，数学分析的发展推动了我们去研究各种曲线和曲面．这只要回想一下，例如曲线正是函数——数学分析的最重要的概念——的几何形象．然而在各种曲线中确实还可以遇到与分析的研究并不相干的曲线．

假如说在古希腊时代已经创立的初等几何学中还不能说到任意的曲线或曲面，那末在解析几何学中已经可以说："每一条曲线都由方程来表示"和"带两个变数 x, y 的任意方程在坐标平面上总表示曲线"．同样地，在坐标系统中曲面由方程 $z=f(x, y)$ 或 $F(x, y, z)=0$ 所给定．确立了几何学和分析学的紧密联系的坐标方法，同时给出了用方程来确定曲线和曲面的方法．

然而解析几何仍然限制于使用代数和初等几何的方法，而且并未深入地进行各种类型的图形的研究．至于任意的曲线和曲面的研究则是一个新的部门——曲线和曲面的理论，它也叫做**微分几何学**．

必须立刻指出的是，微分几何要求所研究的曲线和曲面服从一些普遍的条件，这些条件是为了在研究曲线和曲面时有应用分析学的方法的可能性．然而即使这样，可取的曲线和曲面的种类的多样化在实质上还是没有限制的，因此它们还能以必要的精确性在巨大数量的问题中反映出所研究的现实对象．"微分几何学"这个名称本身表明了这理论所用的方法：它利用微分法作为基本的方法而且首先研究曲线和曲面的"微分的"性质，即曲线和

曲面"在点上"的性质[1]. 例如，曲线在一点处的方向由切线、弯曲程度——曲率(以后将会给出它的确切的定义)等等来刻画. 微分几何学研究曲线和曲面的微小片段的性质，而且只有在较为深入地展开了其理论时才过渡到各种曲线和曲面的"整体的"研究，即在它们的整个延展范围上作研究.

微分几何学的发展与分析学的发展有不可分割的联系. 分析学的基本运算——微分和积分——都有直接的几何意义. 就像在第二章(第一卷)里所已经指出过的那样，对应于求函数 $f(x)$ 的微分的是作曲线

$$y=f(x)$$

的切线. 切线的斜率(即切线对 Ox 轴的倾角的正切)正是函数 $f(x)$ 在对应点处的导数 $f'(x)$ (图 1)，而曲线

$$y=f(x)$$

之"下"的面积，其数值正是这个函数在对应的界限内的积分 $\int_a^b f(x)dx$. 由于在分析学里研究的是任意的函数，所以提到的是任意的曲线或曲面. 在分析学中首先讨论的是研究曲线在平面上的情况的一般步骤：它的上升和下降，弯曲程度的大小，它的凸面所对的方向，改变弯曲之处，等等. 由法国数学家罗皮塔尔在 1695 年出版的第一本分析学教程的名称"用来理解曲线的无穷小分析"，就表明了分析学与曲线理论的紧密联系.

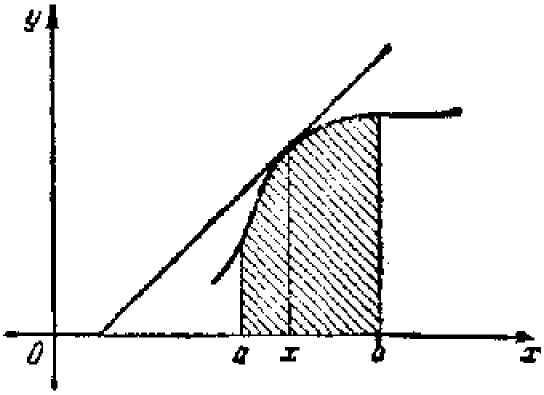


图 1

在十八世纪中叶，当微分法和积分法在牛顿、莱布尼茨和他们

1) 微分几何首先研究曲线和曲面"在点上"的性质(即依赖于已知点的充分小的领域的性质)，并且这些性质由这样一些量来刻划，这些量由在给定曲线或曲面的方程中出现的那些函数(在已知点处)的导数来表达. 就因为这样，所以微分几何要求所研究的曲线和曲面服从一些条件，以便保证应用微分法的可能性：也就是要求曲线或曲面可以用这样的方程来给定，在这些方程中出现的函数具有足够数量的逐次的导数.

的最初的继承者们之后达到了充分的发展时，就显示出了它们在几何学中更进一步的应用的可能性．从那时起已经开始了曲线和曲面理论的本质的发展．对于空间中的曲线和曲面提出了许多问题，它们与平面曲线的情形相仿，但是却无可比拟地具有更复杂和更丰富的内容．这些问题逐渐脱离分析学在几何中的简单应用的范围，而导致独立的曲线和曲面理论的形成．在十八世纪后半有很多数学家参加发展这个理论的原理的工作，他们是克雷洛、欧拉、孟日等等，在其中特别是欧拉，应该被认为是曲面的普遍理论的创立者。关于曲线和曲面理论的第一部独立的著作"分析学在几何中的应用"[1] 是孟日在1795年刊行的书．在这些数学家的研究中，特别是在孟日的书名中，可以明显地看到促使微分几何学发展的因素．那就是力学、物理学、天文学的日益增长的要求，即归根结底技术和工业的要求，而对于这种要求，已有的初等几何是完全不够用的了．

与实践问题有关的还有高斯(1777—1855)关于曲面理论的经典著作．高斯在1827年刊行的著作"关于曲面的一般研究"奠定了作为数学的独立领域的曲面微分几何的基础．在高斯的书中发展了曲面理论的一些普遍的方法和问题，我们将会在§4中谈到它们．高斯特别是从地理制图的要求出发的．地理制图的问题包括如何最精确地在平面上画出地球表面各部分的地图。完全正确的地图在这时是不可能的；尺度比例必须受到地球表面弯曲程度的影响而有所改变．所以就发生了关于寻求最精确的绘图方法的问题．制图学在很古的时代已经开始了，但是它的一般理论的创立还是最近时代的成就，而且没有曲面的一般理论和分析学的一般方法的发展，这种成就是不可能有的。我们注意到，地图制作的困难的数学问题之一是车比雪夫(1821—1894)的研究对象．他所获得的还有关于曲面上的曲线网的重要结果．他还把纯实用的问题

1) 伽斯巴尔·孟日(1746—1828)不仅是一个著名的学者，并且是法国革命的活动家(海军部长，后来是革命军队的大炮和弹药生产的组织者)。他经历了从雅各宾党人到拿破仑帝国拥护者这个法国资产阶级的特有道路。

化成一些这样的研究。

一般的从一个曲面到另一个曲面的映射问题和曲面的变形问题，现在已经是几何学中心部分之一。在这方面的重要结果还是捷尔特（现在的塔尔土）大学的教授闵定格（1806—1885）在1838年得到的。

在上一世纪的后半叶，曲线和曲面的理论已经在它们自己的基础上形成了（假如说的是所谓"古典的微分几何学"，以区别于将在§5中谈到的更新的方面）。已经得到了曲线理论的基本方程——所谓佛锐耐公式。在1853年，闵定格在塔尔土大学的学生彼德逊（1828—1881）的学位论文中获得了和利用了曲面理论的基本的方程；在15年之后这些方程又被意大利数学家科达奇所获得，这些方程通常是以后者命名的。彼德逊在塔尔土读完大学以后，作为中学教师而在莫斯科生活和工作。他虽然没有取得相当于他的杰出科学成就的任何科学院的职位，但是他多少还是莫斯科数学会和从1866年起直到现在都在莫斯科刊行的"数学汇刊"杂志的创立者之一。微分几何学的莫斯科学派就是从彼德逊开始的。

"古典"微分几何学发展的著名的总结是法国几何学家达尔布在他1887—1896年刊行的四卷"一般曲面理论讲义"里作出的。在我们的这个世纪内，古典微分几何继续有所发展，但是在曲线和曲面理论的研究中心，大部分已转向新的方面，在那些方面所研究的图形和其性质的范围还要更为广泛。

§2. 曲线理论

在微分几何学里确定曲线的方法 在分析学和解析几何里我们已知道可用方程来确定曲线。在平面上的直角坐标里，曲线可以用方程

$$y=f(x)$$

或者更普遍的方程

$$F(x, y)=0$$

来给定。然而这种方法只适合于平面曲线（即在平面上的曲线），

可是还必须能用方程来确定空间曲线（即不能含于任何平面内的曲线）. 可以取螺旋线作为这种曲线的例子(图 2).

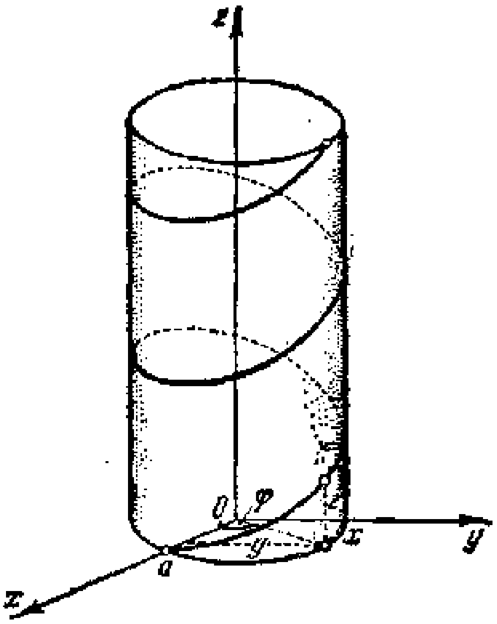


图 2

为了微分几何的目的和在许多别的问题里的需要，最好把曲线看作点的连续运动的轨迹. 当然，已给定的曲线可以从其他来源而得，但是我们总可以设想有着某个点沿着这条曲线而运动.

我们假定在空间中已经取定了一个笛卡儿坐标系. 如果让点 X 以从 $t=a$ 到 $t=b$ 的时间沿着曲线而运动，则这个动点的坐标就是时间的函数 $x(t)$, $y(t)$, $z(t)$. 可以取飞机的飞行或者炮弹的飞驰作为直观的例子. 反之，如果预先给定了函数 $x(t)$, $y(t)$, $z(t)$，则就可以用它们来决定动点 X 的坐标. 随着 t 的改变而运动的点，因而也就画出一条曲线. 因此，空间中的曲线可以用下列形式的三个方程来给定：

$$x=x(t),\ y=y(t),\ z=z(t).$$

同样地，平面上的曲线由两个方程决定：

$$x=x(t),\ y=y(t).$$

这个给定曲线的方法是最普遍的.

让我们来讨论一下螺旋线作为例子. 它在点的螺旋运动下得到，它由绕一条直线——螺旋轴——的等速旋转和沿着这条轴的等速移动合成. 我们取螺旋轴作为 Oz 轴. 设在时刻 $t=0$ 时点在 Ox 轴上. 让我们来找出点的坐标对时间的依赖性. 如果沿 Oz 轴的移动以速度 c 来进行，则显然在这个方面随时间 t 而得的增量是

$$z=ct.$$

如果 φ 是绕 Oz 轴的旋转角而且 a 是从点到轴的距离，则从图 2 可

以看出

$$x=a\cos\varphi,\ y=a\sin\varphi.$$

因为旋转是等速的，所以角 φ 与时间成正比，$\varphi=\omega t$（ω 是旋转的角速度）。因此我们得到

$$x=a\cos\omega t,\ y=a\sin\omega t,\ z=ct.$$

这就是螺旋线的方程；随着 t 的改变，动点以这样的坐标画出一条螺旋曲线.

当然，变数 t，或者像普通所说参数 t，并不必须具有时间的意义.此外，从给定的参数 t 可以转到另一个参数，例如可以用公式 $t=u^3$ 或者一般地 $t=f(u)$ 引进另一个变数 u[1]. 在几何学中，最自然的是取曲线上从某个固定点 A 算起的曲线弧的长度 s 作为参数. 与弧长 s 的每个可能的值对应的是弧 AX 本身. 所以 X 的位置由量 s 完全决定，因而点 X 的坐标就可以表示成弧长 s 的函数，

$$x=x(s),\ y=y(s),\ z=z(s).$$

所有这一些（以及可能的其他一些）给定曲线的方法[2]，都打开了在研究曲线时进行计算的道路.只要用方程表述了曲线，就可以运用数学分析来研究曲线的性质.

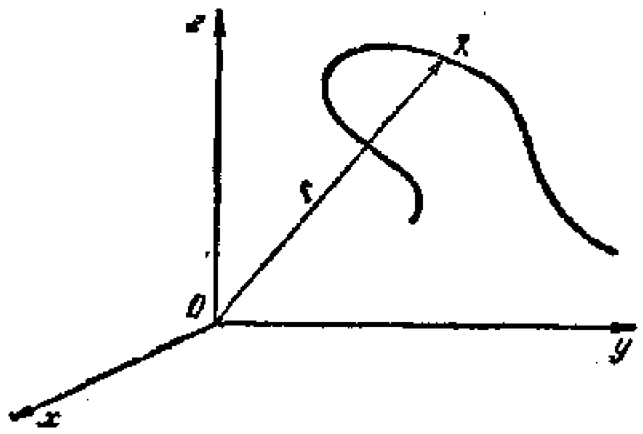


图 3

在微分几何学中，与平面曲线有关的是三个基本概念：长度、切线和曲率. 与空间曲线有关的还有所谓密接平面和挠率. 现在我们就要依次来说明这些概念的意义和价值.

长度 每一个人所已经形成的关于长度的直观看法必须精确

---

1) 这时严格地说，函数 f 必须是单调的.

2) 空间中的曲线还可以作为由方程 $F(x, y, z)=0$ 和 $G(x, y, z)=0$ 决定的两个曲面的交线来给定，即曲线由这两个方程的全体来给定. 在理论的推导中最方便的是用向量来给定曲线，即用从坐标原点引到点 X 的向量 $r=\overrightarrow{OX}$ 来决定这个点的位置. 随着向量 r 的改变，终点 X 画出给定的曲线（图 3）.

化，才能转化成数学曲线的长度的正确定义，这定义引向确定的数值特征，而且被用来以任意的正确程度计算曲线的长度和在谈到长度时作严密的论证．对于所有数学概念都是这样．从不成熟的观念到不变的和正确的定义的过程正是从关于对象的不科学的认识到科学的认识的过程．为了技术和自然科学的需要，概念的正确化终究还是必须的．技术和自然科学的发展要求研究长度、面积和其他一些几何量的性质．

简单的和最合乎需要的长度定义是这样的：曲线的长度是内接于这曲线的折线的长度当折线的顶点在这曲线上无限制地密集时的极限．

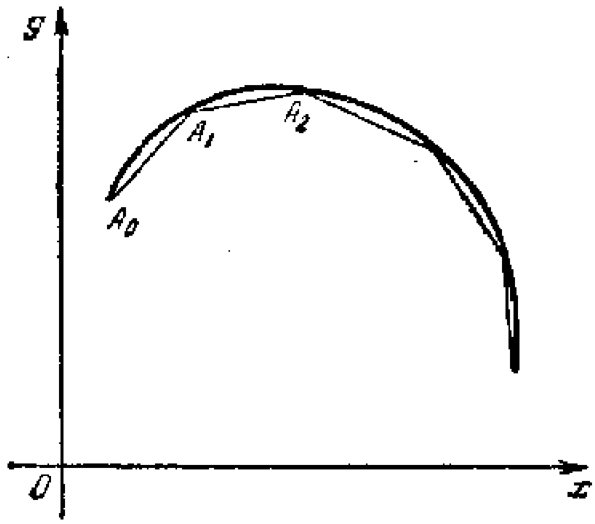


图 4

这样的定义是从自然的测量方法出发的．在曲线上顺次记出点 A_0, A_1, A_2, …（图 4），来测量它们之间的距离．这些距离的总和（这个总和正是内接折线的长度）表示出曲线的近似的长度．为了更正确地决定长度，自然需要更密地选取这些点 A，那时才能更好地照顾到曲线的弯曲．最后，长度的正确的值就被定义为当这些点 A 无限制地密集时的极限值[1]．因此，长度的这个定义正是用愈来愈小的脚步来测量长度的方法的推广．

从长度的定义容易引出计算它的公式，只要曲线是用解析式子给定的．然而我们注意到，数学公式完全不是单为计算用的．他们还代表确立了不同数学量之间关系的定理的简缩写法．这种关系的理论价值可能远远超过了公式的计算价值．举例说，由公式

1) 所说的极限（即长度）的存在，即使对于位在有限制的区域里的曲线，也不是早已知道的．如果曲线非常曲折，则它的长度可以是很大的．可以用数学的方式给出非常"弯曲的"曲线，而使它的任何一段弧都没有有限的长度（内接于这曲线的折线的长度是无界地增大的）．

$$c^2=a^2+b^2$$

表出的毕达哥拉斯定理就完全不是为了计算斜边 c 的平方，而首先是指出直角三角形各边之间的依赖性。

让我们在这里推导出在笛卡儿坐标里由方程 $y=f(x)$ 给定的平面曲线的长度．这时我们假定函数 $f(x)$ 有连续的导数。

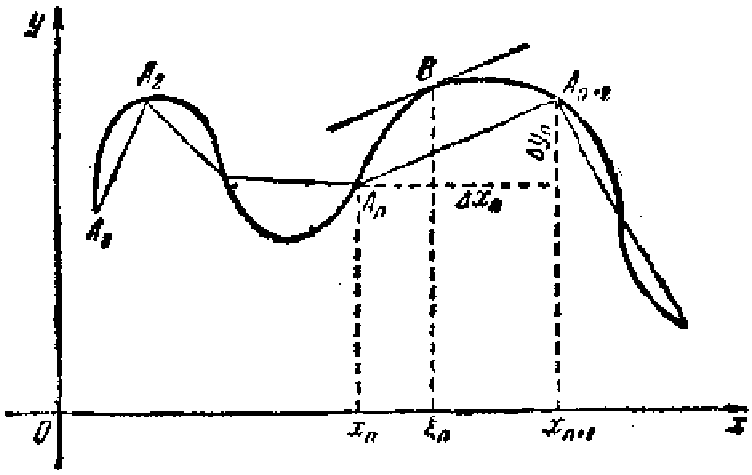


图 5

画出内接于曲线的折线(图 5)．设 A_n, A_{n+1} 是它的两个邻接的顶点，而 x_n, y_n 和 x_{n+1}, y_{n+1} 是这两个顶点的坐标．线段 A_nA_{n+1} 是一个直角三角形的斜边，它的直角边等于

$$\Delta x_n=|x_{n+1}-x_n|,\ \Delta y_n=|y_{n+1}-y_n|.$$

因而，按照毕达哥拉斯定理，

$$A_nA_{n+1}=\sqrt{(\Delta x_n)^2+(\Delta y_n)^2}=\sqrt{1+\left(\frac{\Delta y_n}{\Delta x_n}\right)^2}\,\Delta x_n.$$

容易设想，假如通过点 A_n 和 A_{n+1} 的直线平行地向上或向下移动，则在直线脱离曲线的时刻，它就占有这曲线的某个点 B 处的切线的位置，这就是说，在曲线的片段 A_nA_{n+1} 上至少存在一个点，在这点处的切线的倾角与弦 A_nA_{n+1} 的倾角相同(这个显然的附注容易作严密的证明)。

利用以上所说的，我们把比值 $\frac{\Delta y_n}{\Delta x_n}$ 换成点 B 处的切线的倾角的正切，即把 $\frac{\Delta y_n}{\Delta x_n}$ 换成导数 $y'(\xi_n)$，这里 ξ_n 是点 B 的横坐标．现在这一段的长度就有式

$$\overline{A_nA_{n+1}}=\sqrt{1+y'^2(\xi)}\,\Delta x_n.$$

整个折线的长度是其各段的长度之和. 当用记号 $\sum$ 来缩短加号的写法时, 我们有

$$s_{\text{лом}}=\sum\sqrt{1+y'^2(\xi_n)}\,\Delta x_n.$$

为了得到曲线的长度, 现在只要在量 Δx_n 中的最大者趋向零的条件下过渡到极限,

$$s=\lim_{\Delta x\to 0}\sum\sqrt{1+y'^2(\xi_n)}\,\Delta x_n.$$

但是按照在第二章(第一卷)中所给出的积分的定义, 这个和式的极限正是积分. 那就是说, 这是函数 $\sqrt{1+y'^2}$ 的积分. 因此, 平面曲线的长度由下列公式表出:

$$s=\int_a^b\sqrt{1+y'^2}\,dx,\tag{1}$$

这里积分界限 a 和 b 是 x 在所讨论的曲线弧的端点处的值.

空间曲线长度是对应的, 但是稍有不同的公式, 推导的根据也是同一些东西.

要依照这些公式实地来计算长度, 当然并不总是简单的. 例如圆的长度用公式(1)来计算就很复杂. 但是, 我们已经说过, 公式并不只是用来作计算的; 特别地, 公式(1)对于长度性质的研究, 对于长度与其他几何量的关系等等, 都是很重要的. 在第八章里我们就有用到公式(1)的地方.

切线 平面曲线的切线在第二章(第一卷)里已经讨论过了. 对于空间曲线来说, 切线的意义是完全相似的. 为了在点 A 处决定切线, 在曲线上取与 A 不同的点 X, 而且引割线 AX. 然后让点 X 沿着曲线去接近 A. 如果这时割线 AX 趋向一个极限位置, 则有这个极限位置的直线就叫做曲线在点 A 处的切线[1].

如果区分了曲线的起点和终点, 因而也就区分了它的通行次

---

1) 从第二章图 13 的例子看出, 割线的极限位置可以不存在. 那里所画的曲线 $y=x\sin\dfrac{1}{x}$ 在零的附近上下摆动, 使得当 A 趋向 O 时, 割线 OA 总是从直线 OM 摆动到直线 OL 和反过来从直线 OL 摆动到直线 OM.

序，则就可以说点 A 和 X 中哪一个是第一个，哪一个是第二个.（例如当列车从莫斯科开到海参威时、鄂木斯克显然在伊尔库茨克之前.）由于这个缘故，就可以在割线上用箭头指出从第一个点到第二个点的方向. 这种"有向割线"的极限是"有向切线"(图 6). 切线上的箭头表明沿曲线的运动在通过点 A 的时刻向着哪一个方向. 在点沿着曲线运动时，在每一时刻的运动速度，其方向都沿着它所产生的曲线的切线.

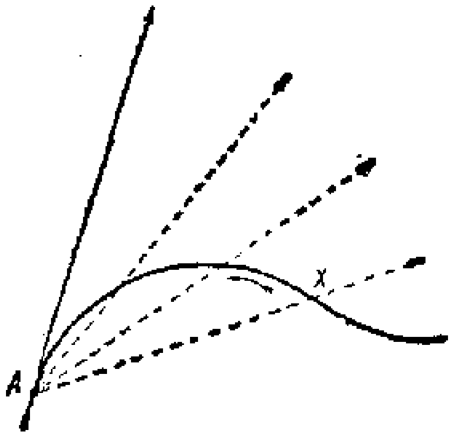


图 6

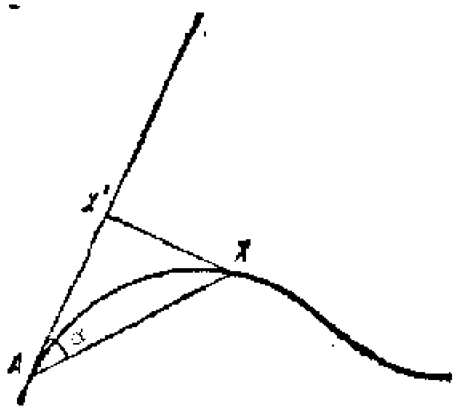


图 7

切线具有一个重要的几何性质：在接近切点处，曲线与这条直线的差异在一定的意义下小于与任何别的直线的差异. 这时曲线上的点到切线的距离与它到切点的距离比起来是很小的. 正确地说，比值 $\frac{XX'}{AX}$(图 7)当 X 趋于 A 时趋向零[1). 因而在曲线的一小段上可以用切线来代替它，所具有的误差与所取线段的大小相比是很小的. 为了简化推导而把曲线的小段换成切线的线段，就是利用切线的这个性质. 在连上极限过程以后，这种方法给出十分正确的结果.

我们有趣地看到，对于不是直线的曲线，即对于没有原来意义的方向的曲线，我们在拿它与直线比较以后，就能在每个点处决定它的方向. 于是方向概念就被推广了，它得到了它原来没有的意

1) 这个断言直接从切线定义本身得出. 实际上，从图 7 看到，$\frac{XX'}{AX}=\sin\alpha$，这里 α 是切线和割线 AX 之间的角. 因而 $\frac{XX'}{AX}$ 随着 α 同时趋向零.

义. 这个新的方向概念反映了沿曲线的运动的现实的本质，即它在每一时刻都有方向，同时连续地改变它的方向.

曲率 要用眼睛来判断道路、细杆或者图上的线段的弯曲程度，并不一定需要数学家. 但是即使对于最简单的力学问题，这种一般的看法也是不够的，而弯曲程度的正确的定性特征是必要的. 这特征的获得明白地表出了关于曲率是曲线方向改变快慢的直观看法中所存在的内容.

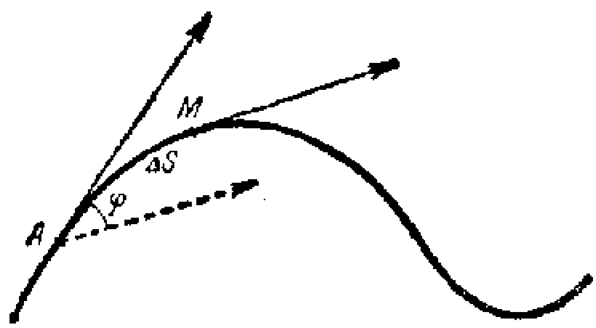


图 8

设 A 是曲线上的点，M 是接近 A 点的(图 8). 在这些点处的切线之间的角表出曲线的从 A 到 M 的片段上的旋转. 我们把这个角记做 φ，旋转的平均速度，更正确地是在长度 Δs 的片段 AM 上以单位长度来算的平均旋转，显然是 $\frac{\varphi}{\Delta s}$. 于是作为曲线在点 A 处的平均旋转速度的曲率，自然就定义为比值 $\frac{\varphi}{\Delta s}$ 当 $M\to A$ 时(换句话说就是当 $\Delta s\to 0$ 时)的极限了. 总之，曲率由下列公式定义:

$$k=\lim_{\Delta s\to 0}\frac{\varphi}{\Delta s}.$$

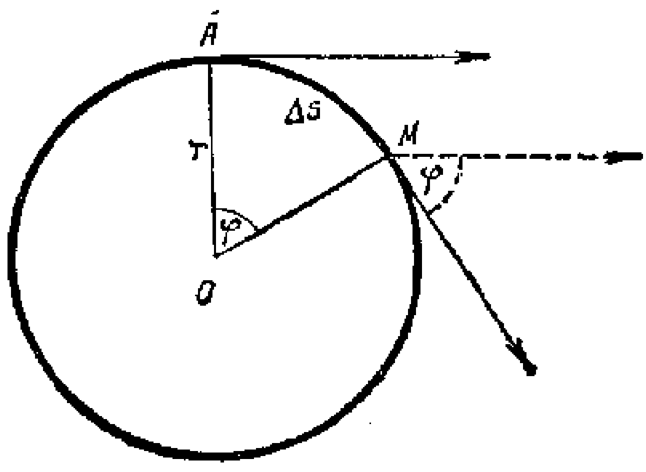


图 9

作为例子，我们来讨论圆的曲率(图 9). 明显地，半径 OA 和 OM 之间的角 φ 和在点 A 和 M 处的切线之间的角 φ 是相等的，这是因为它们的边互相垂直的缘故. 角 φ 所张的弧 AM 有长度 $\Delta s=\varphi r$，于是

$$\frac{\varphi}{\Delta s}=\frac{1}{r}.$$

这说明比值 $\frac{\varphi}{\Delta s}$ 是常数，因而作为这个比值的极限值的圆的曲率，在圆的所有点处都相同而等于半径的倒数[1]。

让我们来导出由方程 $y=f(x)$ 给定的平面曲线的曲率的公式．我们取固定点 N 作为计算长度的起点(图 10)．在点 A 和 M 处的切线之间的角 φ，按绝对值显然等于切线倾角从 A 到 M 的改变量

$$\varphi=|\Delta\alpha|.$$

由于角 α 可能减小，所以我们要取绝对值．

我们所感到兴趣的量

$$k=\lim_{\Delta s\to 0}\frac{\varphi}{\Delta s}=\lim_{\Delta s\to 0}\frac{|\Delta\alpha|}{\Delta s}=\lim_{\Delta x\to 0}\frac{\frac{|\Delta\alpha|}{\Delta x}}{\frac{\Delta s}{\Delta x}}=\frac{|\alpha'|}{s'}.$$

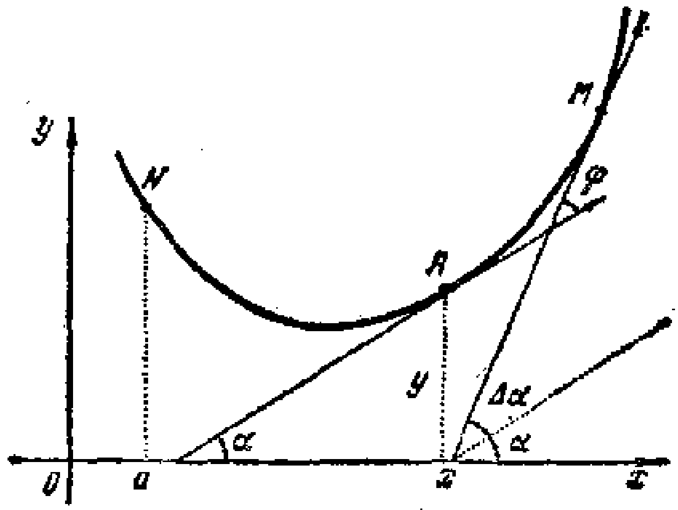


图 10

曲线弧 NA 的长度由积分

$$s=\int_a^x\sqrt{1+y'^2}\,dx$$

表示，因而 $s'=\sqrt{1+y'^2}$．

1) 我们注意到，拿曲线与作为曲率的模范或标准的圆来比较，一般地也可以达到任意曲线的曲率的概念．那就是说：曲率是那样一个圆的半径的倒数，这个圆是在所考虑的点处以最好的方式接近于已知曲线的．

还要来求 α'. 我们知道 $\operatorname{tg}\alpha=y'$; 因而 $\alpha=\operatorname{arc\,tg} y'$. 后面这个等式对 x 求微分，我们得到

$$\alpha'=\frac{1}{1+y'^2}y''.$$

总之，最终地有

$$k=\frac{|\alpha'|}{s'}=\frac{|y''|}{(1+y'^2)^{3/2}}.$$

在普通的分析学教程和微分几何学教程里还导出用别种方法给定曲线时的对应的公式和关于空间曲线的公式.

得到的公式可以用来给出曲率的另一个几何解释，这在很多问题里是有用的. 那就是说，曲线在已知点处的曲率可以用下列公式表出

$$k=\lim_{l\to 0}\frac{2h}{l^2},$$

这里 h 是从曲线上的点到已知点处的切线的距离，l 是切线上从切点到曲线上的点在切线上的投影点这个线段的长度(图 11).

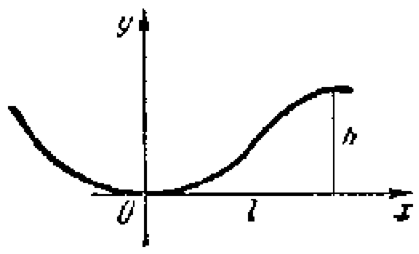


图 11

为了证明，我们取直角坐标系，使坐标原点与曲线的已知点重合，使 Ox 轴与曲线在这个点处相切重合(图 11).(为了简单起见，我们只考虑平面曲线) 于是 $y'=0$, 因而 $k=|y''|$. 按照泰勒公式展开给定曲线的函数 $y=f(x)$ 以后，我们得到: $y=\frac{1}{2}y''x^2+\varepsilon x^2$ (这时考虑了 $y'=0$ 这事实). 这时当 $x\to 0$ 时有 $\varepsilon\to 0$. 由此得出

$$k=|y''|=\lim_{x\to 0}\frac{2|y|}{x^2},$$

而且因为 $|y|=h$, $x^2=l^2$, 所以

$$k=\lim_{l\to 0}\frac{2h}{l^2}.$$

这个公式表明，曲率刻画了曲线离开切线的速度.

我们来谈一下曲线概念与力学问题的一些重要联系.

我们先讨论以下的问题．设有一条柔软的线挂在一个圆柱子上(图 12)，并且这条线处在一个平面上．需要求出在每个点处这线对于柱子的压力，更正确地也就是决定极限

$$p=\lim_{\Delta s\to 0}\frac{P}{\Delta s},\tag{2}$$

这里 P 是长度 Δs 的包含已知点的一段线作用于柱子的力 $\boldsymbol{P}$ 的大小．为了简单起见，我们假定沿整条线的张力 $\boldsymbol{T}$ 的大小 T 是相同的．

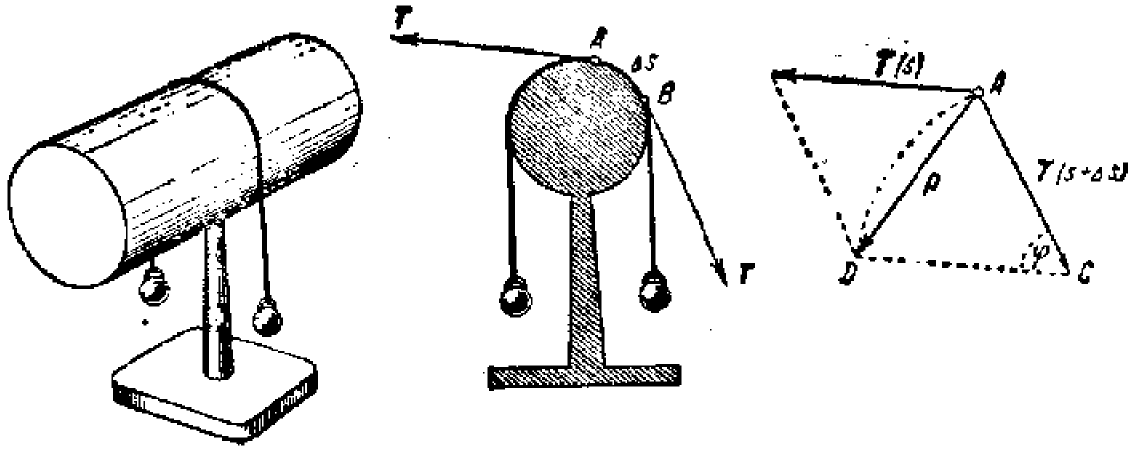


图 12

我们考虑点 A 而且取一段线 AB[1]．在这长度 Δs 的一段线 AB 上，除去柱子的反作用力以外，作用的只是两个外力——端点处的张力，它们大小相等而方向则沿着端点处的切线向着两侧．因而这一段线对柱子的压力 $\boldsymbol{P}$，几何地就等于两个端点处的张力之和．从图 12 可以看到，向量 $\boldsymbol{P}$ 是等腰三角形 CAD 的底边．这个三角形的两腰都等于 T，而顶点 C 处的顶角 φ 则等于切线从 A 到 B 的旋转角．

角 φ 随着 Δs 的减小而减小，而在 $\boldsymbol{P}$ 和点 A 处的切线之间的角则趋向直角．因而压力的方向垂直于切线．

为了求出压力的大小，我们要用到圆的微小的弧按长度接近于它所张的弦这个事实，而且我们就把弦 AD 的长度，即量 P，换

1) 取点 A 是中点的一段线当然更自然些，但是这并不改变最后的结果，反而引起计算上的一些麻烦．

成弧 AD 的长度 $T\varphi$. 于是按公式(2)我们得到

$$p=\lim_{\Delta s\to 0}\frac{P}{\Delta s}=\lim_{\Delta s\to 0}\frac{T\varphi}{\Delta s}=T\lim_{\Delta s\to 0}\frac{\varphi}{\Delta s}=Tk.$$

总之，在每个点处的压力就等于曲率与线上张力的乘积，其方向则垂直于在这个点处的切线.

我们再来讨论另一个问题. 设质点(即很小的物体)沿着平面上的一条曲线以大小不变的速度 v 而运动. 问它在已知点 A 处的加速度是怎样的？按照加速度的定义，它等于速度(在时间 Δt 内)的改变量对时间的增长 Δt 之比值的极限. 这时所取的速度不仅有大小，而且有方向，即我们讨论的是速度向量的改变. 因此，从数学上看，关于加速度的大小的问题就化成求极限

$$w=\lim_{\Delta t\to 0}\frac{|\boldsymbol{v}(t+\Delta t)-\boldsymbol{v}(t)|}{\Delta t},$$

这里 $\boldsymbol{v}(t)$ 是在点 A 处的速度，而 $|\boldsymbol{v}(t+\Delta t)-\boldsymbol{v}(t)|$ 则是表示速度之差的向量的长度. 我们所关心的极限还可以改写成

$$\lim_{\Delta s\to 0}\frac{|-\boldsymbol{v}(t)+\boldsymbol{v}(t+\Delta t)|}{\Delta s}\lim_{\Delta t\to 0}\frac{\Delta s}{\Delta t},$$

这里 Δs 是在时间 Δt 内所通过的弧的长度.

如果注意一下图 13, 而且考虑到速度在每个点处的方向都沿着切线，而它的大小又是常数，那末要以几何方式求出 $-\boldsymbol{v}(t)+$

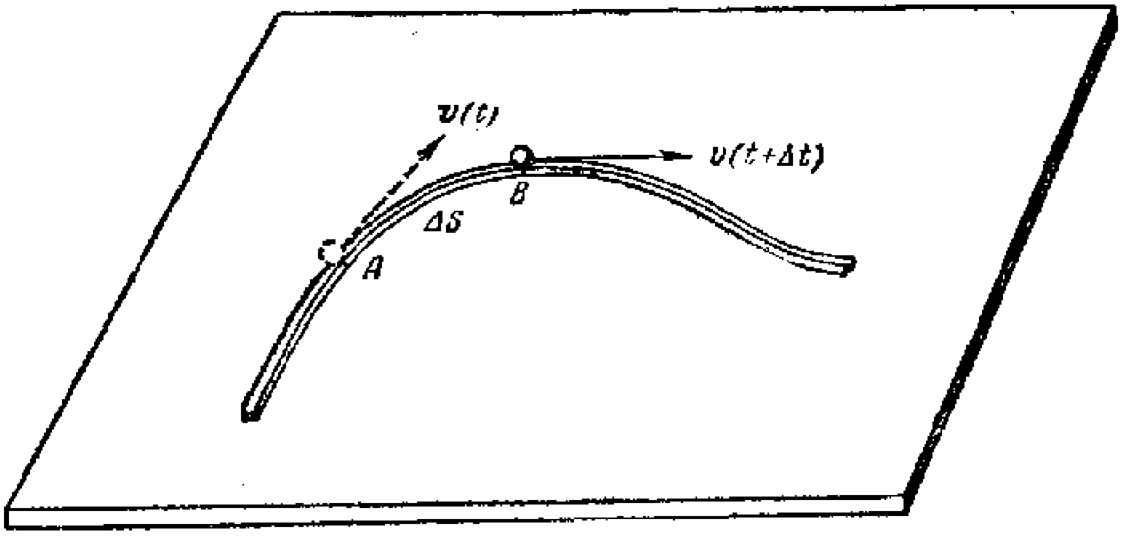


图 13

$v(t+\Delta t)$，正好与上一个问题中求出向量 P 没有任何区别．因而可以利用上一个问题的现成的解，把张力换成速度而写成

$$\lim_{\Delta s\to 0}\frac{|-v(t)+v(t+\Delta t)|}{\Delta s}=vk.$$

此外，$\lim\limits_{\Delta t\to 0}\dfrac{\Delta s}{\Delta t}=v$．因而最终地可以说，物体在沿曲线作均匀运动时所经受的加速度就等于曲率乘上速度平方的乘积

$$w=kv^2, \tag{3}$$

加速度的方向则沿着曲线的法线，即沿着垂直于切线的直线．

引用几何的类比，使我们能够在解决关于加速度的问题时利用关于线的问题的解，这再一次地说明，数学的概念和结论在舍弃了具体特性之后，是如何地促使这些结论富有各种各样应用的可能性呀．

我们还注意到，从力学观点反映了运动方向的改变的曲率，与造成这种改变的力有密切的关系．在等式(3)的两边乘上运动着的质点的质量 m 以后，很容易得到表示这个关系的式子．我们有

$$F_n=mw=v^2mk,$$

这里 F_n 是作用于点上的力在法线方向的分力的大小．

密接平面 虽然空间曲线不在一个平面上，但是通常在它的每个点 A 处可以取一个平面 P，在这个点的邻近，曲线与它的差异小于曲线与任何别的平面的差异．这样的平面叫做曲线在已知点处的密接平面．

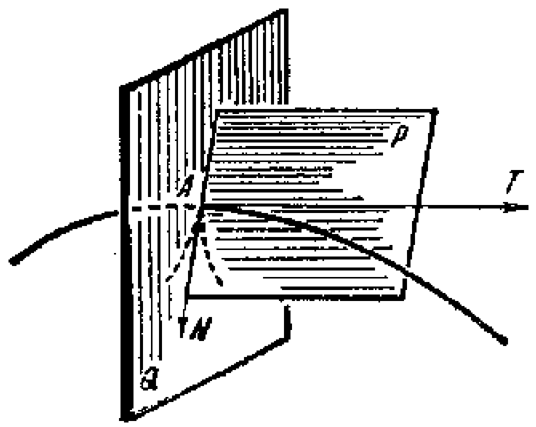


图 14

密接平面自然是这样的平面，它尽可能更紧密地依附于通过点 A 的曲线和已知曲线的切线 T．但是通过点 A 和包含着 A 的直线 T 的平面有很多．为了在所有这些平面中找出曲线与其差异最小的平面，我们试着来考察一下曲线与切线的差异．为

此我们沿着切线来察看曲线，换句话说，我们把已知曲线投射到所谓法平面 Q 上，这是通过 A 而垂直于 T 的平面(图 14)．已知曲线的包含着 A 的片段的投影在平面 Q 上形成一条新的曲线(在图 14 上这条曲线用虚线画出)．这曲线通常在点 A 处有一个尖端．如果得到的曲线在点 A 处有切线 N，则通过 T 和 N 的平面 P 自然就是在 A 邻近最紧密地依附于已知曲线的平面，也就是在点 A 处的密接平面．可以证明，当给定曲线的函数有二阶导数而且曲线在点 A 处的曲率不是零时，密接平面一定存在，而且它的方程可以非常简单地用给定曲线的函数的一阶和二阶导数来表出．

如果说切线的性质可以用来把微小片段的曲线看做直线，而且这样做所造成的误差与曲线片段的长度相比是很小的，那么密接平面的性质使我们能够把微小片段的空间曲线看做平面曲线，用它在密切平面上的投影来代替它，并且这里的误差甚至于与曲线片段的长度平方相比也是微小的．

在空间中垂直于切线的直线是很多的：它们铺满了曲线在已知点处的法平面．在这些直线中可以区别出处在密接平面上的一条直线 N．这条直线叫做曲线的**主法线**．通常它还有固定的方向——向着曲线在密接平面上的投影的凹进的一侧．对于空间曲线说，主法线占有平面曲线的唯一法线的地位．(特别地，如果有一根柱子迫使具有张力 T 的柔软的线保持空间曲线的形状，那么曲线在每个点处对于柱子的压力都等于 Tk 而且沿着主法线的方向．又如果有质点沿着空间曲线以大小不变的速度 v 而运动，那末它的加速度就等于 kv^2 而且方向沿着主法线．)

挠率 当沿着曲线从一个点变到另外一个点时，密接平面的位置自然随之而改变．正像切线的旋转速度由曲率刻画一样，密接平面的旋转速度也由一个新的量——曲线的**挠率**来刻画．这时就像在曲线的情形那样，速度以对于所通过的弧长的比值来计算，即如果 Ψ 是在固定点 A 和邻近它的点 X 处的密接平面之间的角，而 Δs 是弧 AX 的长度，那末在点 A 处的挠率 τ 就定义成

极限[1)]

$$\tau=\lim_{\Delta s\to 0}\frac{\psi}{\Delta s}.$$

挠率的符号依赖于密接平面在沿着曲线的运动下向着哪个方向旋转.

这样一来,可以设想当点沿着曲线运动时,同时随着运动的有密接平面以及画在它上面的切线和主法线,并且切线在每一时刻都向着法线一侧以由曲率决定的速度而转动,密接平面则绕着切线以由挠率决定的速度和方向而转动.

运用微分方程论的最简单的手法,可以证明一个基本定理,大致说来它相当于说有相同曲率和挠率的曲线是相等的.我们来解释一下这个命题.如果从曲线的起点沿着曲线作不同弧长 s 的移动,则依赖于量 s,我们达到曲线上的不同的点,在每个点处都有它独有的曲率 k 和挠率 τ 的值.因而 $k(s)$ 和 $\tau(s)$ 对于每条曲线说都是从曲线起点算起的路程 s 的某些函数.

所说的定理断定,如果两条曲线的曲率和挠率作为弧长的函数是相同的,则这两条曲线是相等的(即其中的一条经过运动可以与另一条重合).这样一来,曲率和挠率作为沿曲线的弧长的函数已经完全决定了这条曲线,只除掉这曲线在空间中的位置,因而可以说曲线的所有性质全都包含在它的长度、曲率和挠率的依赖关系之中.因此这三个概念在研究关于曲线的各种问题时成为一个合理的基础.利用它们还可以获得我们马上就要接触到的曲面理论的最简单的概念.

当然,这一些还不是曲线理论的全部.在其中还可以引出与曲线有关的很多别的概念;研究特殊类型的曲线,曲线族,曲线在曲面上的位置,关于曲线的整体形状的问题,等等.这些问题

---

1) 可以证明螺旋线在其所有的点处有同样的挠率,而且可以从比较曲线与在已知点的邻域里以最好的方式逼近曲线的螺线,来引进任意曲线的挠率的概念.挠率还刻画了曲线对于平面的差异.在与曲率的某种联合下,挠率刻画了曲线离开它的密接平面的速度.

和解决它们的方法几乎与所有的数学分支都有联系．可以利用这个理论的力量来解的问题的范围也是特别丰富和多样化的．

§3. 曲面理论的基本概念

确定曲面的方法 为了用分析学的方法来研究曲面，自然必须能解析地来给定曲面．给定曲面最简单的是用方程

$$z=f(x, y),$$

式中 x, y, z 是曲面上的点的笛卡儿坐标．这时函数 $f(x, y)$ 并不必须对于所有的 x, y 都有定义——它的给定区域可以有各种的结构．例如对于在图15上画出的曲面，$f(x, y)$ 给定在一个圆环的内部．用方程 $z=f(x, y)$ 给定曲面的一些例子，我们在解析几何里已经知道了．例如我们知道，方程 $z=Ax+By+C$ 给定平面；方程 $z=x^2+y^2$ 给定旋转抛物面(图16)．为了应用微分法，函数 $f(x, y)$ 必须有一阶和二阶(有时还需要以后几阶)导数．由这种方程确定的曲面叫做**正则曲面**．几何上表明(虽然不完全正确)曲面是连续地弯曲而没有折断和别的奇异性的．要研究不服从这些条件的曲面，例如有尖点的、有棱的、或有其他奇异性的曲面，需要一些新的研究工具(参看§5)．

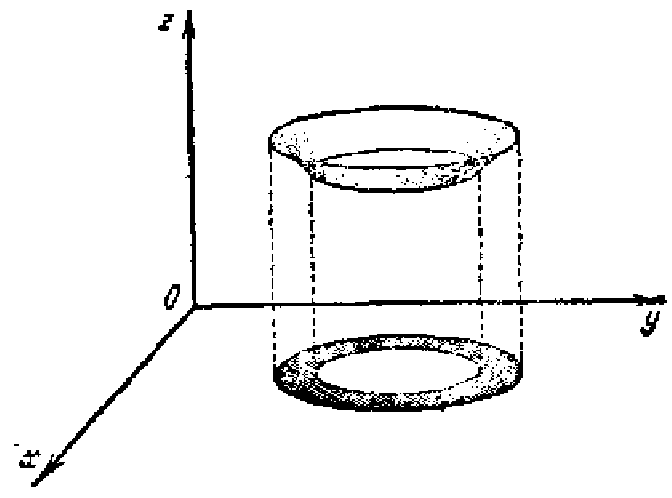


图 15

然而并不是每一个曲面都可以整个地用形状 $z=f(x, y)$ 的方程来表示的，即使是没有奇导性的曲面也是如此。如果与函数 $f(x, y)$ 的给定区域中的每一个值 x, y 对应的是完全确定的 z，则这就说明平行于 Oz 轴的每一条直线都不应该与曲面有多于一个的公共点(图17)。所以即使是非常简单的曲面，如像球面和圆柱面，也不能整个地用形状 $z=f(x, y)$ 的方程来给定。在这些情形里，曲面可以用别种方式，例如用形状 $F(x, y, z)=0$ 的隐方程来给定。在那时候，中心在坐标原点而且有半径 R 的球面有方程

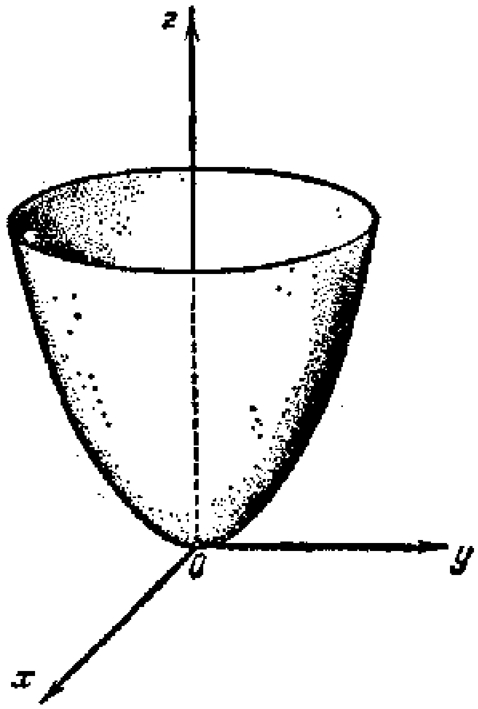


图 16

$$x^2+y^2+z^2=R^2,$$

又方程 $x^2+y^2=r^2$ 给出半径 r 的圆柱面。

当研究的只是曲面的小块时(而在古典的微分几何学里基本

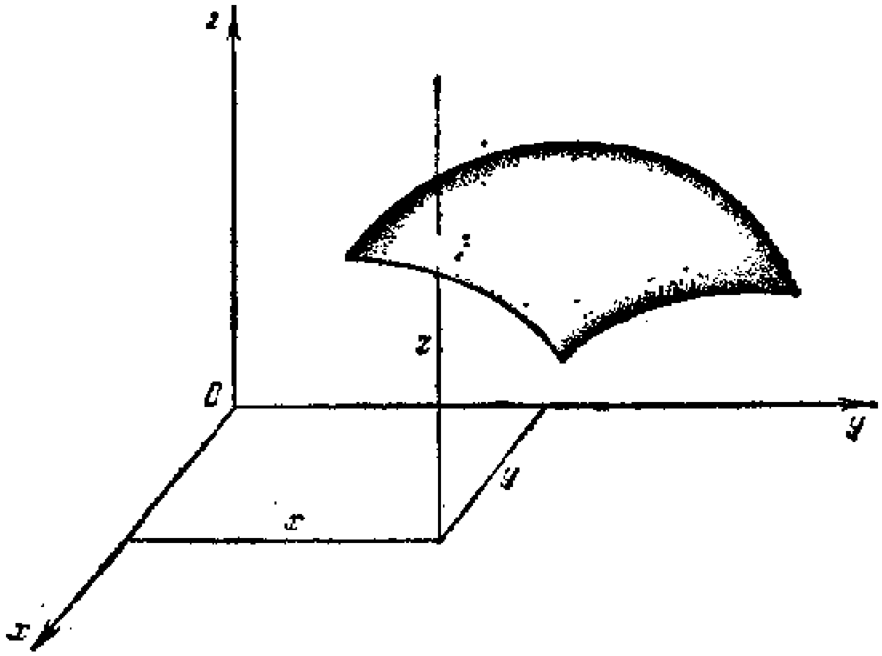


图 17

上就限于这一类的问题)，用方程 $z=f(x, y)$ 给定曲面的方法是完全适合于普遍情形的，因为平滑曲面的每一个充分小的小块都可以表示成这样的形状. 我们就采用这种方法作为基础，至于别的给定曲面的方法则将在 § 4 和 § 5 中谈到.

切平面 正像平滑曲线在每个点处有切线，曲线在这个点的邻域里最接近于这切线一样，许多曲面在它的每个点处都有所谓**切平面**.

下面是切平面的正确的定义. 通过曲面 F 上的点 M 的平面 P 叫做曲面 F 在这个点处的切平面，如果在平面 P 和从 M 引向曲面上任意点 X 的射线 MX 之间的角 α，当点 X 接近点 M 时趋向于零(图 18). 所有与曲面上通过点 M 的曲线相切于点 M 的直线，显然处在切平面上.

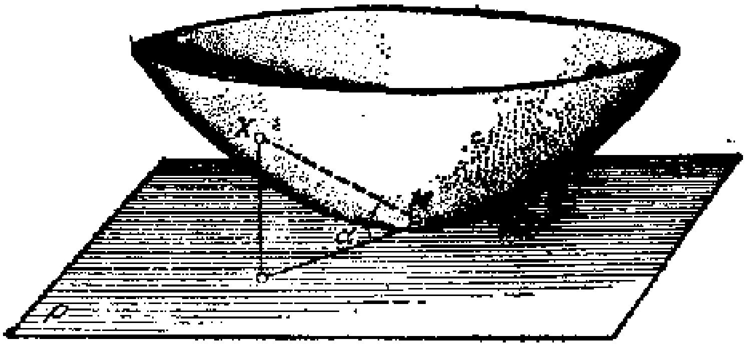


图 18

如果曲面在每个点处都有切平面，而且当从一个点变到另一个点时，切平面的位置是连续地改变的，这个曲面就叫做**平滑曲面**.

在切点邻近，曲面与切平面只有很小的差异，如果点 X 沿曲面接近点 M，则点 X 到切平面的距离与它到点 M 的距离相比是很小的(读者在图 18 上设想点 X 接近于 M，就很容易检验这一点). 这样一来，曲面在点 M 邻近是如何地贴合于切平面呀. 所以在一阶逼近下，曲面的小块，或者说曲面的"元素"，可以换成切平面的小块. 切平面在切点处的垂直线，也是曲面在这个点处的垂直线，它叫做曲面的**法线**.

这种用切平面的小块代替曲面的小块的可能性在很多情形里出现．举例说，光线从曲面上的反射就是像从平面上的反射一样地进行的，这就是说，反射光线的方向由普通的反射定律决定：入射光线和反射光线与曲面的法线处在同一个平面上，而且入射光线和反射光线与法线组成相等的角(图 19)，就像反射是对着切平面而进行似的．同样地，在光线通过曲面而折射时，每一条光线在曲面元素上的折射也按照普通的折射定律，就像这是平面元素一样．在光学器械中所有关于光线的反射和折射的计算，就是根据这一些附注的．再有例如互相接触的刚体在接触点处有公共的切平面．物体通过它们表面的元素而互相接触，而且一个物体在另一个物体上的压力在没有摩擦力时其方向是沿着接触点处的法线的．这个事实在物体不止在一个点处相切时也对，那时在每个接触点处，压力的方向都是沿着对应的法线的．

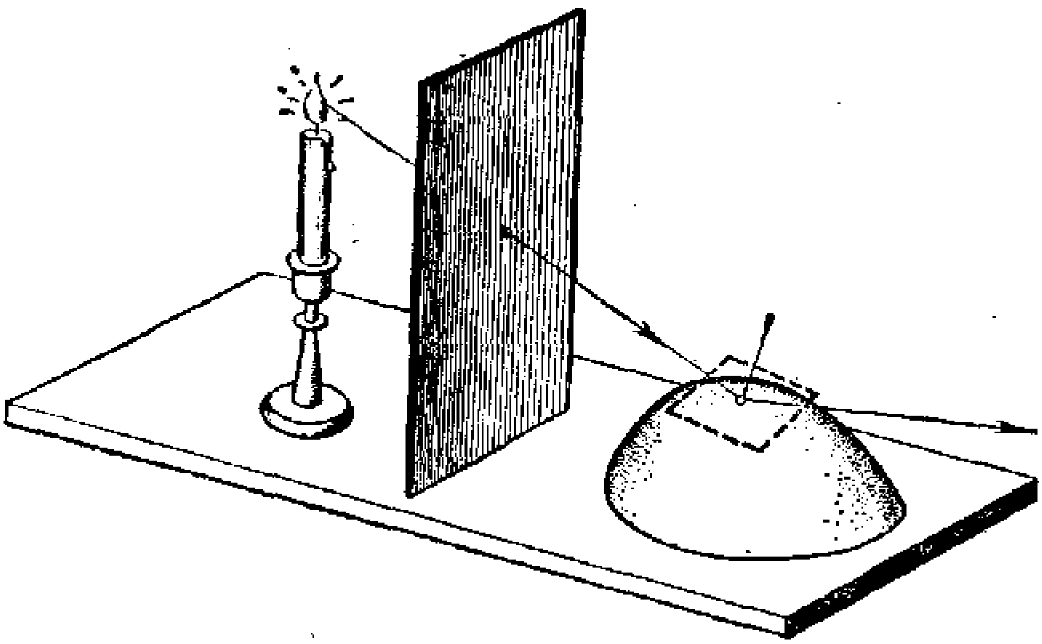


图 19

在用平面的小块来代替曲面元素的基础之上，还可以来定义各种曲面的面积．把曲面分成了小块 $F_1, F_2, \cdots, F_n$，而且把每个小块投射到曲面在这小块的某个点处的切平面上(图 20)．我们得到一些平面区域 $P_1, P_2, \cdots, P_n$．它们的面积之和

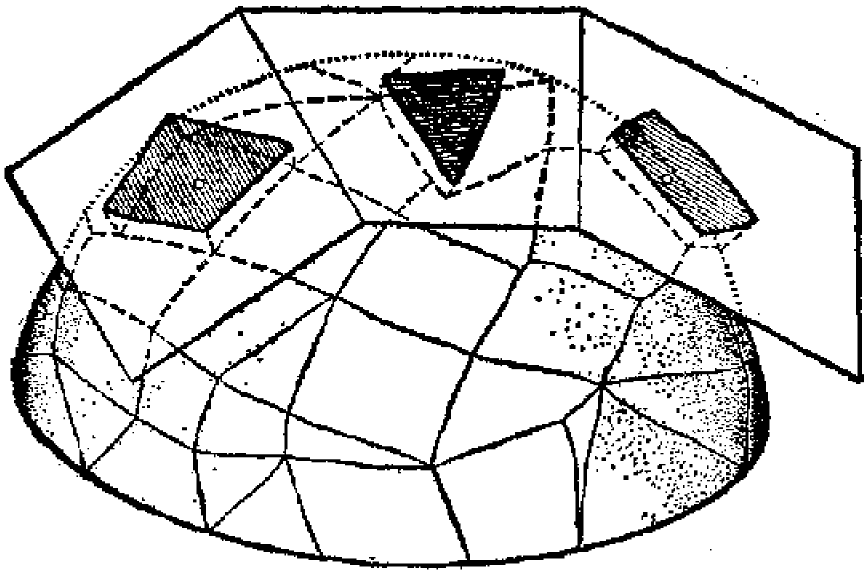


图 20

给出曲面面积的近似值．曲面面积本身就定义成小块 $P_1, P_2, \cdots, P_n$ 的面积之和在曲面分割愈来愈细时的极限[1]．由此可以导出用二重积分表出的正确的面积公式．

从这些附证已经可以看出切平面概念的价值．在很多问题中，曲面元素利用平面来作近似的表示还充分地和必要地照顾到曲面的弯曲性．

曲面上的曲线的曲率　曲面在已知点处的弯曲性由曲面离开其切平面的迅速程度来决定．但是在不同的方向曲面可以以不同的速度离开切平面（例如画在图 21 上的曲面，在 OA 方向就比在 OB 方向更快地离开平面 P）．所以用曲面上向着不同方向的曲线的曲率来决定曲面的弯曲性是很自然的．

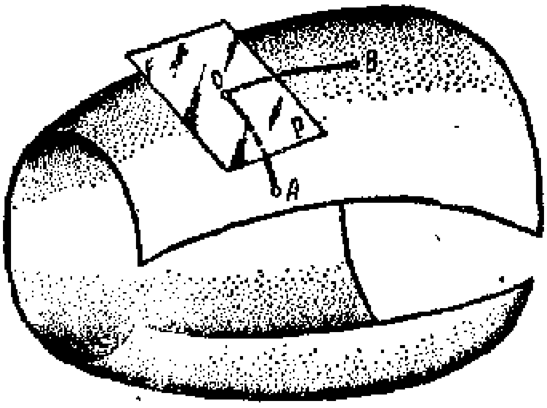


图 21

可以这样地做到．我们通过点 M 引切平面 P 而且取定法线的确定方向(图 22)．我们来讨论通过点 M 处的法线的平面截割曲面而成的曲线；这些曲

1) 在第八章§1里所利用的面积的表达式就是这样地导出的。

线叫做法截线．法截线的曲率带有符号，当它的凹面向着法线一侧时，法截线的曲率取正号，当凹面向着另外一侧时，曲率取负号．例如在图 23 上所画的马鞍状的曲面上，当箭头指出的是曲面的法线的方向时，截线 MA 的曲率被认为是正的，而截线 MB 的曲率则是负的．

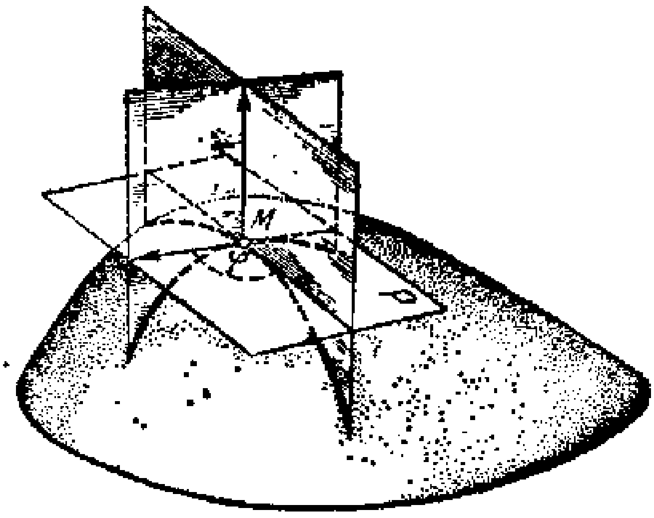


图 22

法截线由它所在的平面与切平面上一条初始的射线所成的角 φ 给定(图 22)．知道了依赖于角 φ 的法截线的曲率 $k(\varphi)$，我们就有了关于曲面在点 M 的邻域里的结构的十分完全的观念．

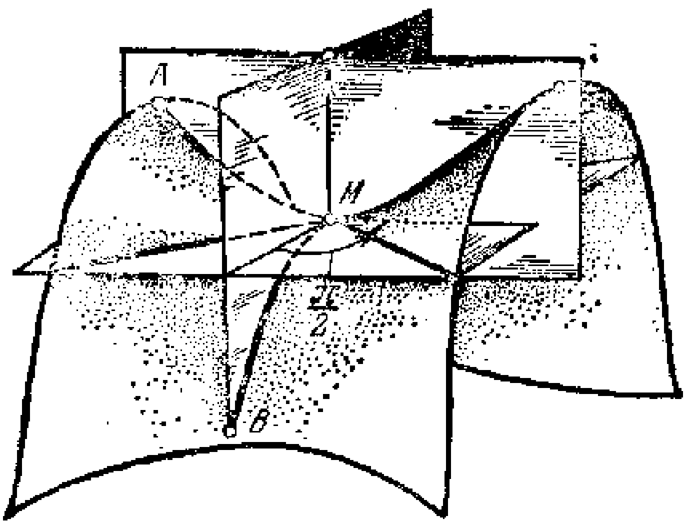


图 23

曲面可以弯曲成各种各样的形式，因而曲率 k 对于角 φ 的依赖关系看来可以是任意的．事实上却并非如此．对于在微分几何里研究的正则曲面，存在着由欧拉发现的简单法则，这法则确定了通过已知点而向着不同方向的法截线的曲率之间的关系．

可以证明，在曲面的每个点处存在着这样两个方向：

(1) 它们互相垂直；

(2) 向着这两个方向的法截线的曲率 k_1, k_2 是所有法截线的曲线中的最大值和最小值[1)]；

(3) 与曲率 k_1 的法截线组成 φ 角的法截线的曲率 $k(\varphi)$ 由下列公式表出：

$$k(\varphi)=k_1\cos^2\varphi+k_2\sin^2\varphi. \tag{4}$$

这样的方向叫做曲面在已知点处的**主方向**，而曲率 k_1, k_2 则叫做曲面在已知点处的**主曲率**.

这个欧拉定理表明，尽管曲面是各种各样的，它们在每个点邻近的结构，以与已知点的变动相比是二阶微小的正确性来考虑，可能有的只是一些完全确定的类型. 事实上，如果量 k_1 和 k_2 有相同的符号，则 $k(\varphi)$ 的符号是不变的，而且曲面在所研究的点邻近有画在图 22 上的形状. 如果 k_1 和 k_2 有不同的符号，例如 $k_1>0$, $k_2<0$, 则法截线的曲率显然要改变符号. 这是因为当 $\varphi=0$ 时曲率 $k=k_1>0$, 而当 $\varphi=\frac{\pi}{2}$ 时我们有 $k=k_2<0$.

从关于 $k(\varphi)$ 的公式(4)不难肯定，当 φ 从 0 变到 π 时，$k(\varphi)$ 的符号改变两次[2)]，因此，在曲面的所讨论的点邻近，曲面有马鞍的形状(图 23).

当数 k_1 和 k_2 中有一个是零时，那么曲率总保持同一符号，但是对于 φ 的一个值它成为零. 例如在柱面(图 24)上的每一个点处就是如此. 在一般情形中，曲面在这种点邻近有接近于柱面的形状.

最后当 $k_1=k_2=0$ 时，所有的法截线都有零曲率. 在这种点邻近曲面特别"密切地"依附于切平面. 因而这种点叫做**平面化的点**. 在图 25 上给出了这种点的一个例子(点 M). 曲面在平面化

1) 特别当 $k_1=k_2$ 时，所有法截线的曲率都相同(例如在球的情形).

2) 通过简单的计算可以证明，当 $\varphi=\operatorname{arctg}\sqrt{-\frac{k_1}{k_2}}$ 和 $\varphi=\pi-\operatorname{arctg}\sqrt{-\frac{k_1}{k_2}}$ 时，$k(\varphi)=k_1\cos^2\varphi+k_2\sin^2\varphi$ 成为零，第一种情形符号从正变到负，第二种情形从负变到正.

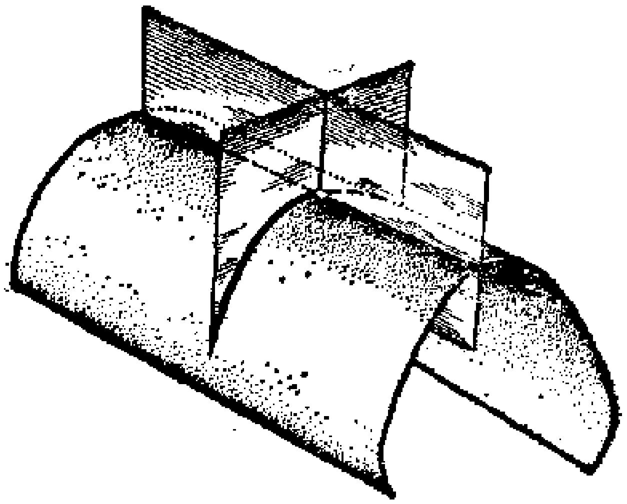


图 24

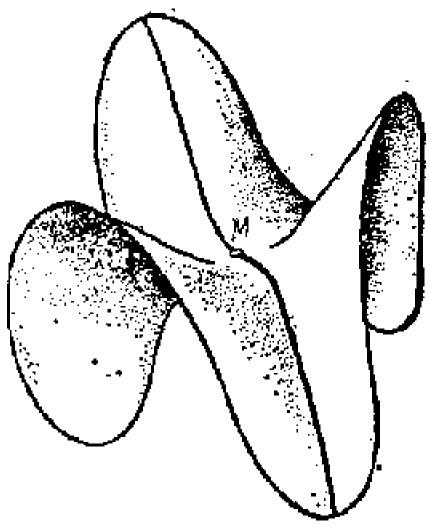


图 25

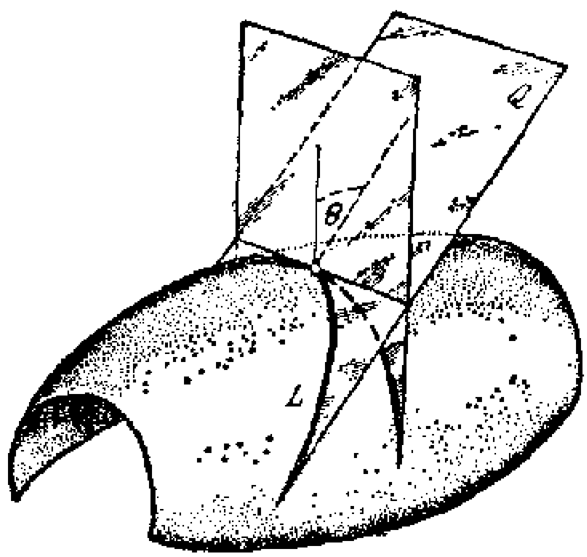


图 26

的点邻近的性质可以是非常复杂的.

现在让我们来讨论曲面与任意不通过法截线的平面 Q 的截线(图 26). 梅涅[1] 曾经证明过，这种线 L 的曲率 k_L 与向着同一个方向(即切平面与平面 Q 的交线的方向)的法截线的曲率 k_H 有着简单的关系. 这个关系由下列公式表出:

$$k_L=\frac{k_H}{\cos\theta},$$

1) 梅涅 (1754—1793)是法国的数学家而且是孟日的学生，是革命军队的将领，他因为在战争中受伤而死亡.

这里 θ 是法线和平面 Q 之间角的（在球面的例子里可以特别直观地看出这个公式的真正意义）.

最后，曲面上以平面 Q 作为密接平面的任意曲线的曲率，可以证明是重合于曲面与平面 Q 的交线的曲率的.

总之，在知道了 k_1 和 k_2 以后，曲面上的任意曲线的曲率就由它的切线的方向以及它的密接平面和曲面法线之间的角决定. 因此，曲面在已知点处的弯曲特性由两个数 k_1 和 k_2 决定. 这两个数的绝对值等于两个互相垂直的法截线的曲率，它们的符号则表明对应法截线的凹面方向对于曲面法线的取定方向的关系.

让我们来证明上面所引进的欧拉定理和梅涅定理.

1. 在证明欧拉定理时我们利用下面的引理. 如果函数 $f(x, y)$ 在已知点处有连续的二阶导数，则坐标轴可以旋转这样一个角 α，使得在新坐标里，混合导数 $f_{x'y'}$ 在这个点处等于零[1]. 我们知道，经过转轴，新变数 x', y' 与旧变数 x, y 的关系由下列公式表示：

$$x=x'\cos\alpha-y'\sin\alpha;\quad y=x'\sin\alpha+y'\cos\alpha$$

[参看第三章(第一卷)§ 7]. 为了证明这个引理，我们注意到

$$\frac{\partial x}{\partial x'}=\cos\alpha,\quad \frac{\partial y}{\partial x'}=\sin\alpha,\quad \frac{\partial x}{\partial y'}=-\sin\alpha,\quad \frac{\partial y}{\partial y'}=\cos\alpha.$$

现在可以按照复合函数的微分法则来计算导数 $f_{x'y'}$. 经过计算我们得到

$$f_{x'y'}=f_{xy}\cos 2\alpha+\frac{1}{2}(f_{yy}-f_{xx})\sin 2\alpha,$$

由此容易推出，当

$$\operatorname{ctg} 2\alpha=\frac{1}{2}\,\frac{f_{xx}-f_{yy}}{f_{xy}}$$

时确实有 $f_{x'y'}=0$.

现在让我们来讨论由方程 $z=f(x, y)$ 给定的曲面 F，并且我们取所研究的 M 作为坐标原点，而在切平面 P 上选取坐标轴 Ox 和 Oy，使得 $f_{xy}(0, 0)=0$. 我们在平面 P 上取与 Ox 轴组成角 φ 的任意直线，来讨论沿着这条直线方向的法截线 L（图 27）. 按照在 § 2 中所引入的公式，曲线 L 在点 M 处的带有符号的曲率等于

$$k_L=\lim_{\xi\to 0}\frac{2f(x, y)}{\xi^2}.$$

1) 我们利用了缩写的偏导数记号，例如把 $\frac{\partial f}{\partial x}$ 写成 f_x 把 $\frac{\partial^2 f}{\partial y^2}$ 写成 f_{yy} 等等.

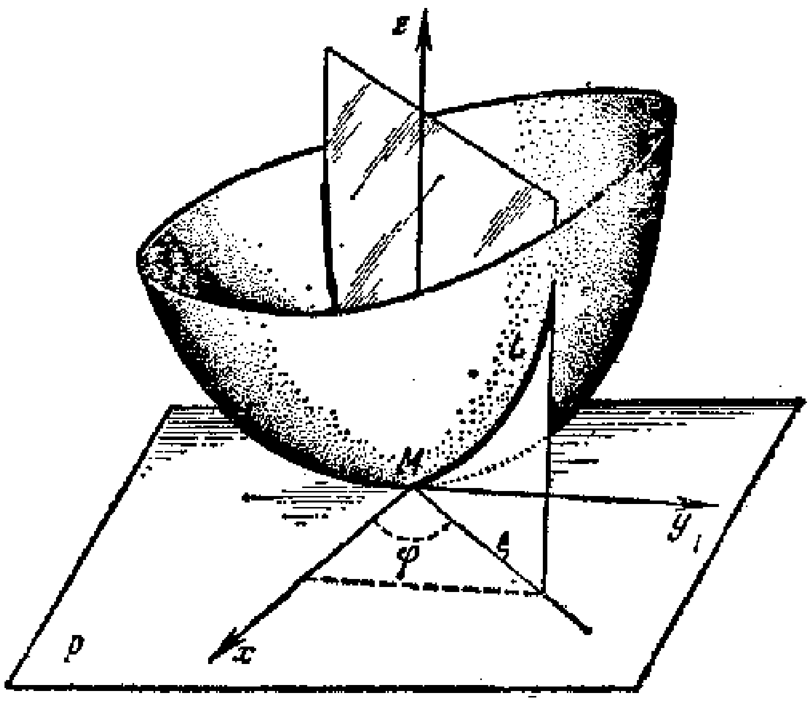


图 27

这里 $f(x, y)$ 是 L 上的点到所取直线的带有符号的距离. 按照泰勒公式 (第二章 § 9)展开 $f(x, y)$ 而且注意到 $f_x(0, 0)=f_y(0, 0)=0$ (因为 Ox 轴和 Oy 轴都在切平面上), 我们得到

$$f(x, y)=\frac{1}{2}(f_{xx}x^2+f_{yy}y^2)+\varepsilon(x^2+y^2),$$

这里当 $x\to 0$, $y\to 0$ 时有 $\varepsilon\to 0$. 对于 L 上的点, 我们有 $x=\xi\cos\varphi$, $y=\xi\sin\varphi$, $\xi^2=x^2+y^2$ (图 27), 因此我们得到

$$\begin{aligned} k_L&=\lim_{\xi\to 0}\frac{f_{xx}\xi^2\cos^2\varphi+f_{yy}\xi^2\sin^2\varphi+2\varepsilon\xi^2}{\xi^2}\\ &=f_{xx}\cos^2\varphi+f_{yy}\sin^2\varphi. \end{aligned}$$

令 $\varphi=0$, $\varphi=\frac{\pi}{2}$, 我们肯定 f_{xx}, f_{yy} 就是 Ox 轴和 Oy 轴方向的法截线的曲率 k_1, k_2. 所以得到的公式就是欧拉公式, $k=k_1\cos^2\varphi+k_2\sin^2\varphi$ (至于说 k_1 和 k_2 是最大的和最小的曲率则从这个公式就可以推出).

2. 为了证明梅涅定理, 我们来讨论法截线 L_H 和另一条截线 L, 截线 L 的平面从截线 L_H 的平面旋转一个角 θ 而得到, 就像在图 28 上所画的一样. 我们还让 Ox 轴和 Oy 轴都在切平面上, 使得 Ox 轴在坐标原点处切于曲线 L_H 和 L. 曲线 L 上有坐标 x, y, $f(x, y)$ 的点 X 到 Ox 轴的距离 $h(x, y)$ 显然等于(见图28),

$$h(x, y)=\frac{|f(x, y)|}{\cos\theta}.$$

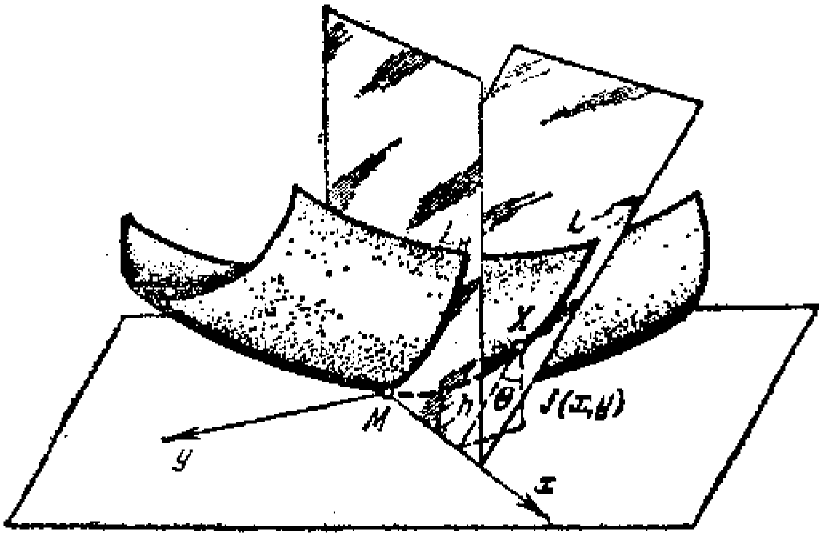


图 28

利用泰勒公式，我们照以下的方式来变换曲线 L 的曲率 k_L 的表达式：

$$k_L=\lim_{x\to0}\frac{2h(x,\ y)}{x^2}=\lim_{x\to0}2\frac{|f(x,\ y)|}{x^2\cos\theta}$$

$$=\lim_{x\to0}\frac{|f_{xx}x^2+2f_{xy}xy+f_{yy}y^2+2\varepsilon(x^2+y^2)|}{x^2\cos\theta},\qquad(5)$$

并且当 $x, y\to0$ 时有 $\varepsilon\to0$. 因为 Ox 轴切于曲线 L，所以显然有 $\lim\limits_{x\to0}\frac{y}{x}=0$. 因而在公式(5)中过渡到极限，我们得到

$$k_L=\frac{|f_{xx}|}{\cos\theta}.$$

但是在我们所选取的坐标系里，曲线 L_H 有方程 $z=f(x,0), y=0$，对于这条曲线有 $|k_H|=|f_{xx}|$. 因而 $k_L=\frac{|k_H|}{\cos\theta}$ 梅涅定理也就证明了.

平均曲率 在曲面理论的很多问题里，起重要作用的不是主曲率本身，而是与它们有依赖关系的量：所谓曲面在已知点处的平均曲率和高斯曲率(或称总曲率). 我们就要详细地来谈一下它们.

曲面在已知点处的平均曲率是指主曲率的和的一半，即

$$K_{\text{cp}}=\frac{1}{2}(k_1+k_2).$$

为了引入利用这个概念的例子，我们来讨论下面的力学问题.

假定在某个物体 F 的表面上紧紧地绷上了一块有弹性的(设是橡皮的)薄膜．问在物体 F 表面的每个点处薄膜的压力有多大？

点 M 处的压力是由薄膜作用于包含点 M 的曲面小块的单位面积上的力来量出的；正确地说，"在点 M 处"的压力是作为这样一个比值的极限来量出的，那就是所说的力对小块面积的比值当小块向着点 M 缩小时的极限．

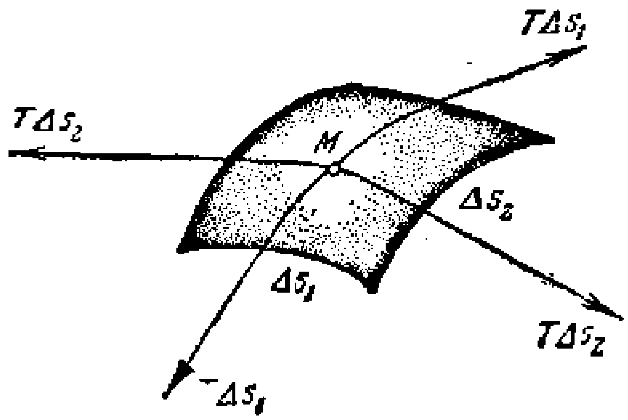


图 29

我们在曲面上点 M 周围取一个不大的曲边矩形，它的各边有长度 Δs_1 和 Δs_2 而且分别向着垂直于点 M 处的第一和第二主方向(图 29)[1]．在矩形的每条边上作用着与边长和薄膜的张力 T 成正比的力(根据薄膜张开时的假定了的均匀性)．所以在垂直于第一主方向的一双边上作用着近似地等于 $T\Delta s_1$ 的力而且其方向沿着曲面的切线．同样地，作用于矩形另一双边的是等于 $T\Delta s_2$ 的力．为了求出在点 M 处的压力，必须拿这四个力的合力被矩形的近似地等于 $\Delta s_1\Delta s_2$ 的面积除，然后在 Δs_1, $\Delta s_2\to 0$ 时过渡到极限．让我们单独地先把前两个力加起来，再用 $\Delta s_1\Delta s_2$ 去除它们的合力．

如果从侧面来观察这个曲边矩形(图 30)，则可以看出这两个力的方向沿着第一法截线的切线，而且它们所在的两点之间的距离正好等于 Δs_2．所以我们所关心的极限的计算问题正是与在 § 2 里已经解决过的线在柱上的压力问题相同的问题．利用原来的结果，就得到我们所关心的极限等于 k_1T，这里 k_1 是第一法截线的曲率．同样地讨论了另外两个力，我们最后引出公式

$$P_M=T(k_1+k_2)=2TK_{\text{cp}}.$$

1) 我们的论证并不严密，然而在估计了这里所许可的误差以后，可以更严密地确立同一个结果．

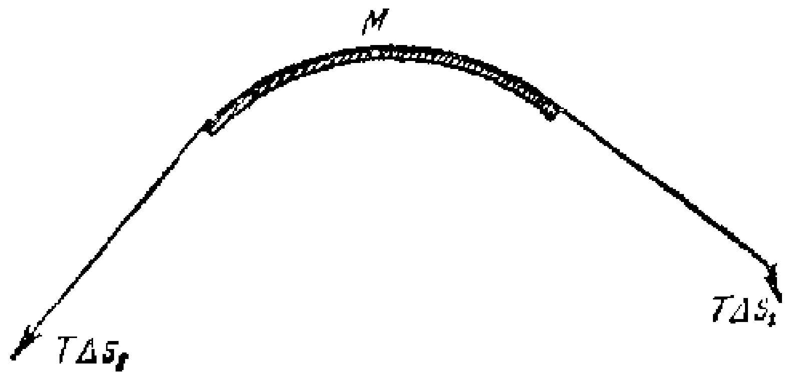


图 30

得到的结果有很多重要的推论．我们来讨论一个例子．

已知液体的表面薄膜在所有方向都有同样的一个表面张力．当液体边界有弯曲的形状时，按照上面所说的，就是这个张力产生了与边界曲面在已知点处的平均曲率成比例的表面薄膜在液体上的压力．

由于这个原因，在非常小的一滴里必须分布巨大的压力，才能阻止它分成一些更小的小滴．在蒸汽冷却时，通常是在灰尘和带电微粒周围形成水滴的．在真空里，小量重新冷却的蒸汽是很难形成水滴的．而如果让具有很大速度的微粒通过这些蒸汽，拨动分子的游离状态，则环绕着微粒所通过的路径，离子在一瞬间形成蒸汽的小滴而成为微粒的可见的痕迹．（在核子物理学中就是根据这一点来建立应用很广的威尔逊雾室的，它是被利用来观察各个带电微粒的运动的．）

由于液体能向各个方面均匀地传送压力，液体小滴在没有别的压力来源时应该具有这样一种形状，使得它的表面在所有的点处都有同样的平均曲率．在所谓柏拉图实验里有两种同样比重的液体，其中一种液体凝聚成块悬浮在另一种液体的内部而呈平衡状态．可以认为悬浮液体受到的只是由张在其边缘的曲面所造成的压力[1]．这时可以证明，"悬浮的"液体总具有球的形状．实验结果引出一种想法，有定常平均曲率的每一个闭曲面是球面．这个

1) 压力随深度而增加一节可以忽略不计，因为它对于两种液体说是相同的(由于它们比重相同)．因而在它们的各部分的边缘，由深度所造成的内外补充压力是彼此平衡的．

定理确实是对的，但是它的严密的数学证明却非常困难.

还可以从另一侧面来达到这个的目的．由于液体的表面薄膜趋于缩小，而液体的体积是不能改变的，自然就要求悬浮液体在一定的体积下具有最小的表面．可以证明，具有所说性质的立体也是球体.

我们所得到的在薄膜的侧面压力和它的平均曲率之间的关系，还可以应用到张在一定周界上的肥皂薄膜的形状问题上．在这时候由于薄膜的沿着曲面法线方向的压力并无任何支撑物的反作用力来平衡(这时薄膜根本就没有套在任何支撑物上)，所以这种压力应该等于零，而对于所求的曲面我们就有条件

$$K_{cp}=0. \tag{6}$$

利用平均曲率的解析表示，从这个条件得到一个微分方程，于是问题就化成在所求曲面通过已知周界的条件下来解这个微分方程[1]．这个困难问题牵涉到很多的研究.

求张在已知周界上的极小面积的曲面的问题也引出同一个方程(6)．从物理的观点看来这种一致是很自然的，因为薄膜趋于缩小，而且唯有当其表面在已知条件下达到可能的极小面积时，才是稳定的平衡状态．由于这后一个问题，零平均曲率的曲面叫做极小曲面.

极小曲面的数学研究之所以引起大家的注意，部分原因是在肥皂薄膜的实验中出现了大量的各种各样的极小曲面．在图31上就画着张在不同周界上的两种肥皂薄膜的图象.

高斯曲率　曲面在已知点处的高斯曲率是指主曲率的乘积

$$K=k_1k_2.$$

高斯曲率的符号决定了曲面在所讨论的点邻近的结构的特性．当 $K>0$ 时曲面有碗的形状(k_1 和 k_2 同号)，当 $K<0$ 时，k_1 和 k_2 不同号，曲面有马鞍的形状．我们以前说过的曲面结构的其他情形相当于高斯曲率等于零的情形．高斯曲率的绝对值给出的

1) 对于由方程 $z=z(x, y)$ 给定的曲面，方程(6)可以化成

$$(1+{z'_y}^2)z''_{xx}-2z'_xz'_yz''_{xy}+(1+{z'_x}^2)z''_{yy}=0.$$

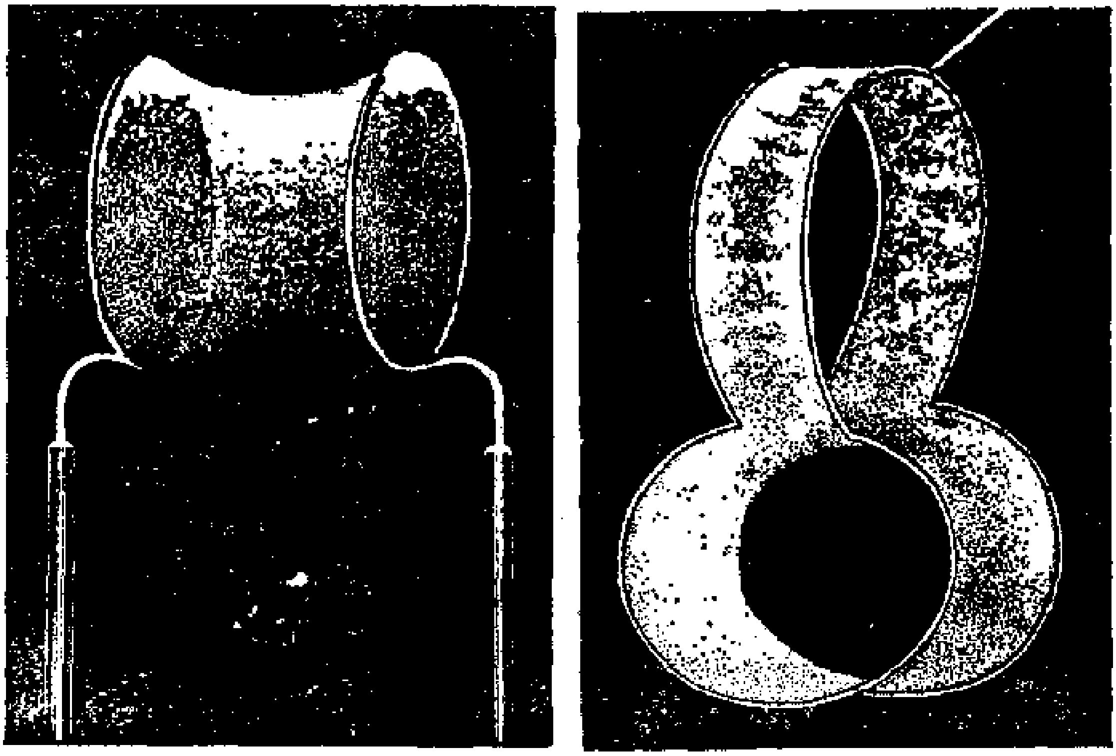


图 31

是从各个方向的曲率分布抽象出来的曲面的一般弯曲程度的观念. 假如转向高斯曲率的另一个定义, 不运用关于曲面上的曲线的研究成果, 以上事实就成为特别明显的了.

让我们来讨论曲面 F 的不大的一块 G, 它包含着点 M 于其内部, 在这一块曲面的每个点处作出曲面的法线.

如果从某一个点画出这些法线, 则它们形成一个立体角(图32). 当小块 G 愈大而且曲面在这小块上弯曲得愈利害时, 这个立体角的量就愈大. 因而小块的弯曲程度可以用法线所形成的立体角的量对于小块 G 本身的面积的比值来刻画; 曲面在已知点处的弯曲程度自然就用这个比值当 G 收缩成点 M 时的极限来量出了[1]. 可以证明, 这个极限就等于在点 M 处的高斯曲率的绝对值.

决定高斯曲率在曲面理论中的地位的是以下的重要性质. 设

1) 为了量出立体角, 可以取立体角的顶点为中心作出一个单位球面. 球面与立体角相交而成的区域的面积就可以取作立体角的量(图32).

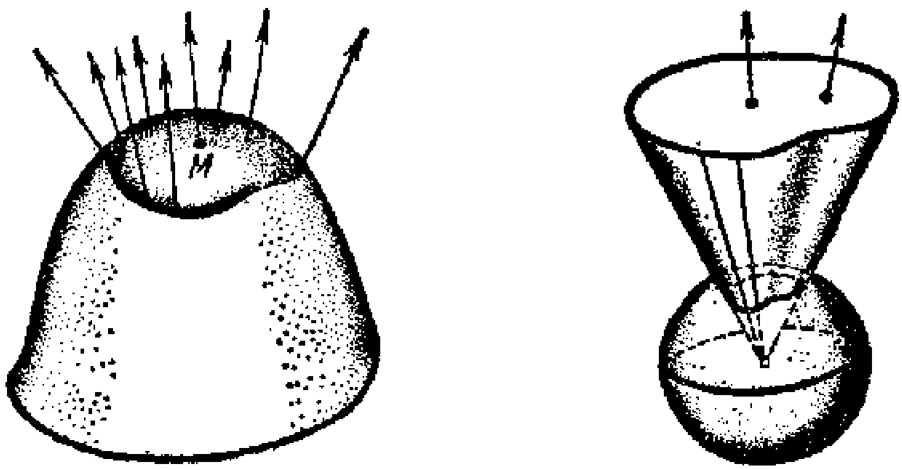


图 32

想所讨论的曲面由可以弯曲但是无伸缩性的材料(例如薄洋铁皮)制成．它的一小块可以经受弯曲，使它改变形状而不把制成它的材料拉长或折断．这时主曲率将要改变，但是正像高斯所证明的，在每个点处的主曲率的乘积 k_1k_2 却保留不变．曲面理论中的这个重要结果表明，高斯曲率不同的曲面具有很大的差别性，即使允许曲面经过可能的弯曲——无伸缩的变形，也不能使这样两个曲面彼此叠合．举例说，球面的小块随便怎么弯曲也不能"铺"到平面上或放置在有另一个半径的球面上．

我们讨论了曲面理论的一些基本的概念．至于这个理论所运用的方法，则在一开头就已经说过，它们首先包括分析学的应用，特别是微分方程理论的应用．我们在证明欧拉定理和梅涅定理时，已经举出了利用分析学的最简单的例子．我们应该指出，为了解决更复杂的问题，还需要把曲面理论的问题化成分析学的问题的特殊方法．这种方法的基础是引入所谓曲纹坐标，这是高斯在其与下一节将要讲述的问题有关的工作中首先作了深入发展的．

§4. 内蕴几何和曲面的弯曲变形

内蕴几何 上面已经说过，曲面的弯曲变形是指它的这样的变形，它保留了曲面上曲线的长度不变．举例说，卷成筒状的纸片

从几何观点看来就是平面小块的弯曲变形．事实上，这时曲面确实没有伸展，而且画在纸上的所有曲线的长度在卷起纸片时也没有改变．保留不变的还有另外一些与曲面有关的几何量，例如曲面上的图形的面积．曲面在弯曲变形下不改变的所有性质，就组成曲面的所谓**内蕴几何**的对象．

这是什么样的一些性质呢？显然，在任意的弯曲变形下可以保留的只是那样一些性质，它们只以有限步的计算依赖于曲线的长度，即它们可以用在曲面上所产生的测量方法来确定．弯曲变形是保留曲线长度的任意变形，而经过任何弯曲变形都不改变的每一个性质都可以这样或那样地通过长度来决定．大家说，内蕴几何简单地就是**曲面上的几何**．“内蕴几何”这名词的本身的意义是说，研究的只是曲面本身的内在的性质，而不依赖于曲面在空间中是怎么样弯曲的[1]．举例说，如果我们在纸片上用直线段连结两个点，然后弯曲这张纸(图 33)，则线段就变成一条曲线，然而它是曲面上连结两个已知点的最短曲线这个性质仍然保留；因此它属于内蕴几何．反之，这条曲线的曲率依赖于纸片的弯曲程度，因此已经不归于内蕴几何了．

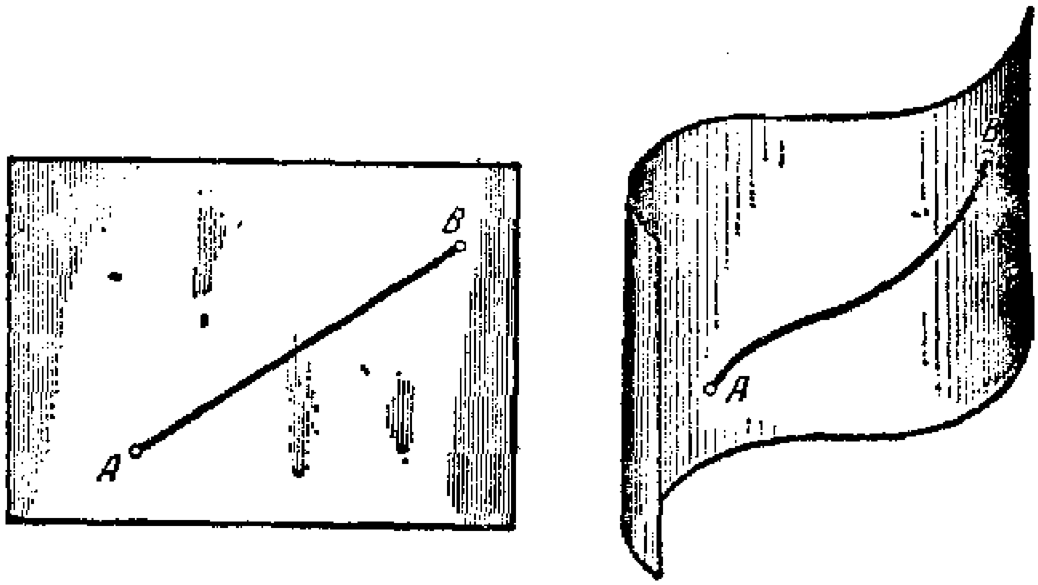


图 33

1) 我们要指出的是，内蕴几何的概念推广空间的数学概念，因而它在现代物理学中占有非常重要的地位，在第十九章(第三卷)里将会详细地说到这一点．

一般地说，由于平面几何的结论不牵涉到包容这个平面的空间的性质，平面几何的全部定理都属于从平面的弯曲变形所得到的任意曲面的内蕴几何．可以说，平面几何是平面的内蕴几何．

内蕴几何的另一个大家知道的例子是球面几何，在测量地球表面时我们实质上就要用到它．这个例子特别适宜于说明内蕴几何概念的本质．事实是，由于地球有很大的半径而把直接看到的一块地面理解成平的，因而在测量很大的距离时而观察到的与平面几何的差异就出现在我们面前，并非作为地球表面在空间中的弯曲的结果，而是作为由地球表面本身的几何性质所表示的“地面几何”所特有的法则．

应该指出，研究内蕴几何的观念本身当在高斯那儿产生时就是与测量学和地图制图学有关的．这两种实用科学实质上都与地球表面的内蕴几何有联系．地图制图学处理的特别是当把一部分地球表面画到平面上时比例尺所受到的歪曲，因此也就要处理地球表面的内蕴几何与平面几何的差异．同理可以想像其他曲面的内蕴几何：设想在已知曲面上生活着某种微小的生物，在这种生物的视界之内曲面看来像是平的（我们知道，任何平滑曲面的充分小的一片与切平面只有很小的差别）；那么这种生物就不会注意到曲面是在空间中弯曲的，只有在测量较大的距离时，他们才会肯定他们的几何学服从另外一些法则，这些法则正是相当于他们所生存的曲面的内蕴几何的．至于这些法则确实随曲面的不同而有差别，则可以用以下的论断来肯定．在曲面上取一个点 O，我们来讨论这样的曲线 L，在曲面上量得的这曲线上的任意点到点 O 的距离（即连接这点与点 O 的最短曲线的长度）都等于常数 r（图 34）．曲线 L 从内蕴几何的观点看来正是半径 r 的圆周．用来表示曲线 L 的长度 $s(r)$ 依赖于 r 的公式是属于已知曲面的内蕴几何的．然而这种依赖关系却可以是各种各样的：例如在平面上 $s(r)=2\pi r$；在半径 R 的球面上不难算出是 $s(r)=2\pi R\sin\frac{r}{R}$；在图 35 所画的曲面上，从 r 的某个值开始，具有已知中心 O 的圆周的长度

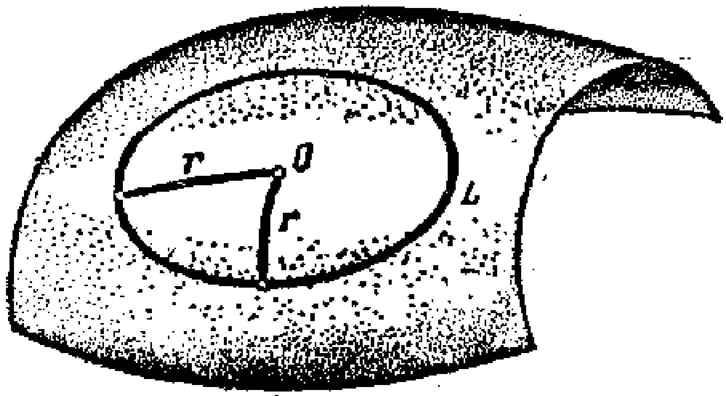


图 34

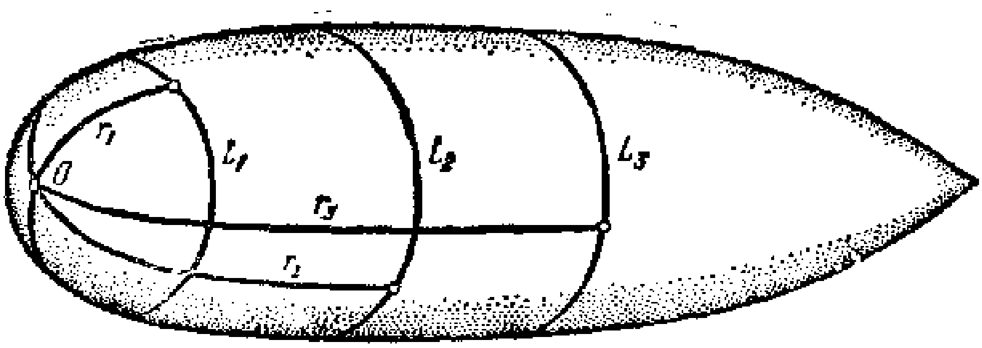


图 35

变得完全不依赖于 r，而后来又变成递减的．因此，以上所说的各种曲面具有不同的内蕴几何．

内蕴几何的基本概念 为了说明内蕴几何的概念和定理的范围，让我们来看一下作为平面的内蕴几何的平面几何．平面几何的对象是平面上的图形和它们的性质，这些性质通常是以长度、角度、面积等等基本几何量之间的关系表出的．当然，角和面积属于平面的内蕴几何的严格的根据是在于：可以证明角和面积都能用长度来表出．这是因为，只要知道已知角所在的三角形各边的长度，就可以算出这个角；三角形的面积也可以按它的各边来计算，而任意多角形的面积则可以把它分成三角形来计算．

把平面几何看作平面的内蕴几何，并不必需限制在中学平面几何的范围．相反地，它可以展开得非常远，加进新的问题，只要所引用的概念以有限次测量长度的计算作为基础就成．例如在平面几何里可以顺次地引入曲线长度的概念，曲边图形面积的概念

等等；它们全都属于平面的内蕴几何。

在任意曲面的内蕴几何里可以引进同样一些概念．曲线的长度在这时是原始的概念；至于角和面积情况稍有些复杂．假如已知曲面的内蕴几何与平面几何有差别，那么我们就不能按普通的公式用长度来决定角和面积．然而，我们曾经说过，曲面在已知点邻近与其切平面很少差异．确切地说，下面的断言成立：假如把包含已知点 M 的曲面小片投射到这点处的切平面上，那么在曲面上量得的两点之间的距离与它们的投影之间的距离的差，比起它们到点 M 的距离来是高于二阶的无穷小量．所以在决定属于曲面的已知点的几何量时，假如是用不高于二阶的无穷小量为主的极限过程得到这些几何量，则就可以把曲面的小片换成它在切平面上的投影．这时从切平面上量得的量对于曲面来说就是内蕴几何的量．这种把曲面小片看作平面的可能性是定义内蕴几何的所有概念的基础．

作为例子，我们来讨论角和面积的定义．依据一般的原则，曲面上曲线之间的角定义为它们在切平面上的投影之间的角（图36）．显然，用这种方法定义的角与曲线的切线之间的角重合．在§3中给出过的面积的定义基于同样的原则．最后，为了刻画曲线在曲面"内部"的弯曲需要引进测地曲率的概念；"测地曲率"这个名称是从地球表面上的测量而来的．曲线在已知点处的测地曲率

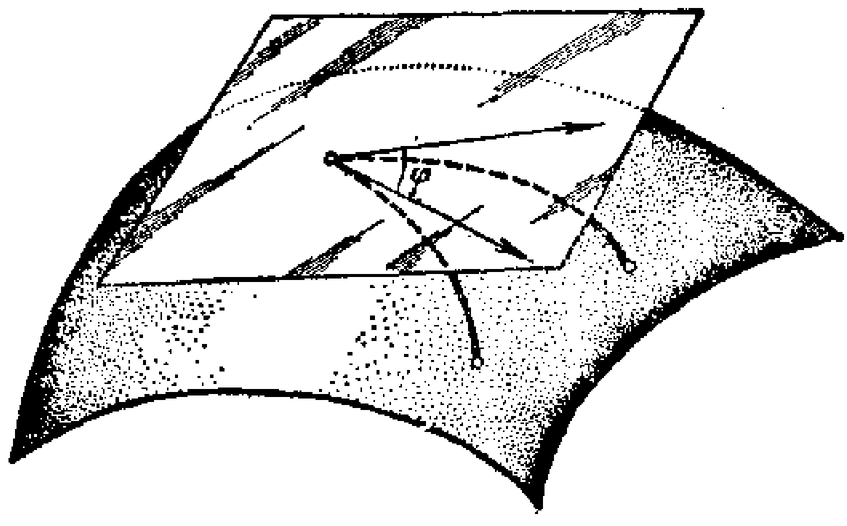


图 36

定义为它在切平面上的投影的曲率(图 37).

这样一来，我们断定平面几何的基本概念都可以引进任意曲面的内蕴几何中去.

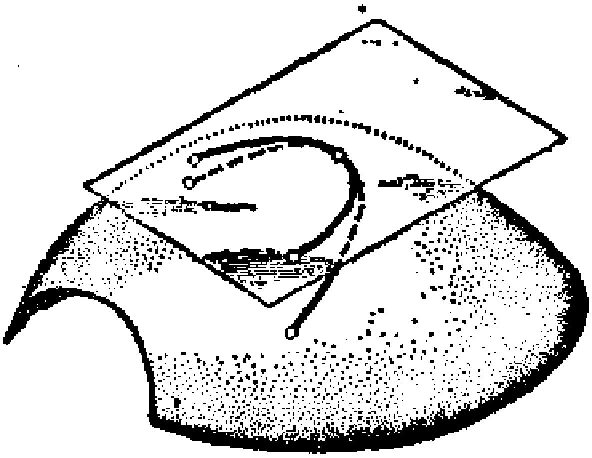


图 37

还很容易在任意曲面上定义与平面上的基本图形相仿的图形. 譬如说，我们已经谈到过圆，它就可以像在平面上一样来定义. 还可以定义线段的同类物——"最短线"作为曲面上连接两个已知点的曲线中的最短者. 然后,自然可以定义三角形(作为以三条最短线为界的图形)、多角形等等. 然而所有这些图形和几何量的性质都依赖于曲面本身，因而在这意义下就存在着无穷多种不同的内蕴几何. 但是内蕴几何作为曲面理论的特殊部分，主要研究的是在任意曲面的内蕴几何里都成立的那些一般的法则，而且在这时需要说明这些法则如何地通过刻画已知曲面的几何量而表示出来.

我们曾经指出过，曲面的最重要的特征之一——它的高斯曲率——在弯曲变形下不变，即只依赖于曲面的内蕴几何. 原来高斯曲率就主要方面还刻画了曲面在已知点邻近的内蕴几何与平面几何的差异程度. 作为例子，我们在曲面上讨论以已知点 O 为中心而且有极小的半径 r 的圆. 在平面上这种圆的周长 $s(r)$ 由公式 $s(r)=2\pi r$ 表示. 在不同于平面的曲面上，圆周长度对半径的依赖关系是另一个样子的；这时可以证明，当 r 很小时，$s(r)$ 与 $2\pi r$ 的差别主要依赖于圆心处的高斯曲率 K，那就是说，

$$s(r)=2\pi r-\frac{\pi}{3}Kr^3+\varepsilon r^3,$$

这里当 $r\to 0$ 时 $\varepsilon\to 0$. 换句话说，当 r 很小时，圆周长度可以按照普通的公式来计算，误差是三阶微小，并且这个误差(以高于三阶的精确性)与高斯曲率成正比. 特别地，如果 $K>0$，则微小半

径的圆周长度比平面上同样半径的圆周长度小些，如果 $K<0$，则相反地要来得大些．不过后面这个结论不难直观地看出，在曲面的具有正曲率的点邻近曲面的形状像"碗"，曲面上的圆周就短些；而在具有负曲率的点邻近，圆周沿着"马鞍子"而成波浪形，因而有些拉长(图 38)．

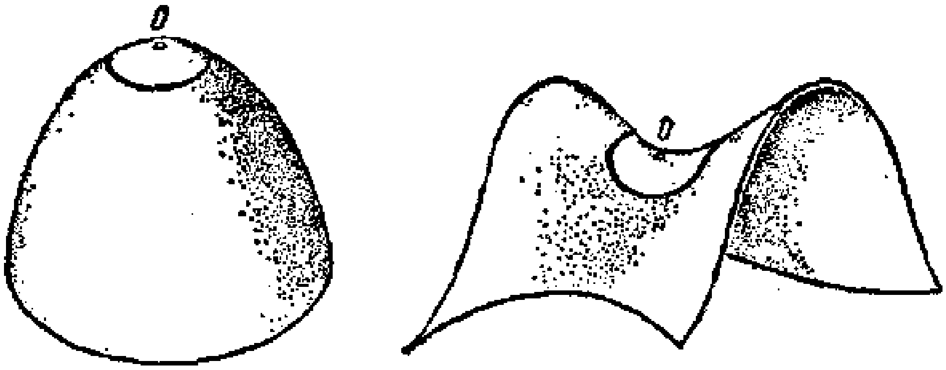


图 38

从以上所引的定理推出，具有变动的高斯曲率的曲面是几何地不均匀的，它的内蕴几何的性质随着点的改变而改变．如果说内蕴几何问题的特征使它接近于平面几何，那么上述的不均匀性就成为它与平面几何的很大的原则性的区别所在．譬如说，在平面上任何三角形的各角之和都等于两个直角；而在任意的曲面上，由最短线组成的三角形的各角之和是不确定的，即使知道了三角形所在的曲面和指出了三角形的"大小"(例如各边的长度)．然而，如果知道了在这三角形的每个点处的高斯曲率 K，则它的各角 α, β, γ 之和就可以按下列公式来计算：

$$\alpha+\beta+\gamma=\pi+\iint K\,d\sigma,$$

这里积分是对三角形的面积来求的．作为特例，这个公式包括了关于平面上和单位球面上的三角形各角之和的熟知的定理．在平面上，$K=0$ 和 $\alpha+\beta+\gamma=\pi$，而在单位球面上，$K=1$ 和 $\alpha+\beta+\gamma=\pi+S$，这里 S 是球面三角形的面积．

可以证明，高斯曲率为零的曲面的任何充分小的片段都可以弯曲(或者更常说成展开)到平面上，因为它有与平面相同的内蕴

几何．这种曲面叫做**可展曲面**．而如果高斯曲率接近于零，则虽然曲面已经不能展开到平面上，但是它的内蕴几何与平面几何只有很小的差别．这再一次表明，高斯曲率表示曲面的内蕴几何与平面几何的差异的程度．

测地曲线　在内蕴几何里起着直线的作用的是所谓测地曲线，它也常常简单地叫做“测地线”．

平面上的直线可以定义为由彼此有一部分重叠的线段组成的线．同理可以定义测地线，只是线段换成了最短线．换句话说，测地线是曲面上的这样的曲线，它的每一个充分小的弧都是最短线．并非每一条测地线就整体来看都是最短线，关于这一点可以以球面为例，球面上的每一个大弧都是测地线，但是只有不超过半圆周的大圆弧才是最短线．我们还看到，测地线甚至可以是闭曲线．

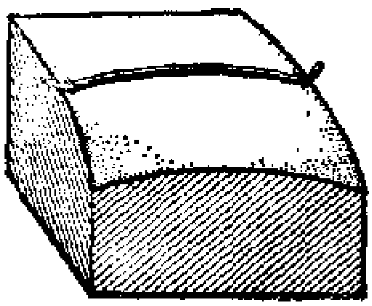


图　39

为了说明测地线的一些重要的性质，我们来讨论下面的力学模型[1]．设在曲面 F 上绷着一条两端固定的拉紧的橡皮线（图39）[2]．当橡皮线有最短的长度时，它就处在稳定状态，因为它每一改变位置都需要拉长，因而只有在外力的作用下才可能发生．这说明沿最短线放置的线处在平衡状态．为了平衡，必须在每一段线上的弹力都与曲面的向着法线方向的反作用力相平衡．（我们认为曲面是平滑的，而且在线和曲面之间没有摩擦力．）但是在§2里曾经确定过，由柱上拉紧的线所产生的压力，其方向向着沿这线的曲线的主法线的方向．所以我们导出了这样的结果：测地线在每个点处的主法线与曲面的法线同方向．逆定理也成立：在正则曲面上的具有所说性质的任何曲线都是测地线．

测地线的上述性质可以用来导出下面的重要事实：如果质点

1）我们早已说过，我们下面的论述并不要求严格证明测地线的性质．我们的目的只是说明其中的一些重要性质罢了．

2）拉紧的线只能绷在凸曲面上，因此为了不造成例外，最好设想曲面是两层的，而线则绷在它的两层之间．

沿曲面运动时，除掉曲面的反作用力以外不受任何其他力的作用，则它的轨迹就是一条测地线. 事实上，我们从§2知道，点的沿脊法线的加速度的方向与轨迹的主法线方向相同，而由于作用于点上的唯一的力是曲面的反作用力，轨迹的主法线就要与曲面的法线重合，所以根据上述的逆定理，轨迹是测地线. 测地线的最后这个性质还大大地加深了它与直线的相仿性. 正像自由的点在惯性作用下沿着直线运动一样，拘束在曲面上的点在没有外力时沿着测地线而运动[1].

从测地线的同一个性质还可以推得下面的定理：如果两个曲面沿着一条曲线彼此相切，而这曲线在其中一个曲面上是测地线，那么这曲线在另一个曲面上也是测地线. 事实上，因为在这曲线的每个点处两个曲面有公共的切平面，所以它们在这些点处有公共的法线，而因为在其中的一个曲面上曲线是测地线，这法线又与曲线的主法线重合. 因此，在第二个曲面上曲线也是测地线.

从这个结果还推得测地线的两个直观的性质. 其一，假如把有弹性的长方形薄片(例如钢皮尺)沿着它的中分线紧密地贴在曲面上，那么它就与这曲面沿着测地线而相切. (实际上，由于接触曲线在尺上是测地线，所以它在曲面上也是测地线.)其二，如果有某个曲面在平面上沿着一条直线而滚动，那么这直线在曲面上的痕迹是一条测地线[2]. 这两个例子都可以举圆柱为例来说明和利用把平面长条的中分线贴到圆柱的实验来肯定(图 40)，这时中分线的位置或者沿着母线，或者沿着圆周，或者沿着螺旋线(不难证明，圆柱面只有这几种类型的测地线). 假如让圆柱在用白粉画出直线的平面上滚动，那么在圆柱上出现的也是这同一些曲线.

测地线与平面上直线的相仿性还可以补上一个重要的性质，这个性质常常被取作测地线的定义. 那就是说，平面上的直线可以定义为零曲率的曲线，而曲面上的测地线则可以定义为测地曲

---

1) 这里所谓"外力"我们是指除曲面反作用力外的所有的力.

2) 这个命题实质上与上一个命题等价，因为曲面在平面上以所说的方式滚动就相当于把一条平的带子贴到曲面上.

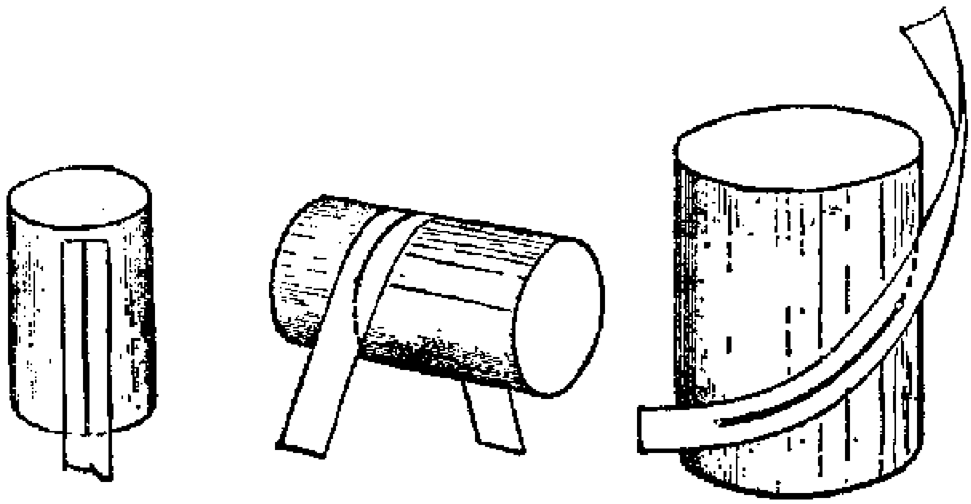


图 40

率为零的曲线.（我们记得，当考虑曲线在曲面的被研究的点处的切平面上的投影时，曲线在这点处的测地曲率就是这投影在同一点处的曲率；参看图 37.）测地线的这个定义重合于原来定义的自然性可以用下面的想法来加以说明.如果在某条曲线的每个点处切平面上的投影的曲率等于零，那么曲线在离开切线时主要就向着曲面法线的方向，因而曲线的主法线也就正好向着曲面法线的方向，而曲线就显得是原来意义下的测地线．反过来，如果曲线是测地线，则它的主法线(因而也就是曲线与切线的差异的主要部分)方向向着曲面法线的方向，所以在把曲线投射到切平面上时就得到这样一条曲线，它与切线的差异大大地小于原来的曲线与切线的差异，而且所得的投影的曲率就等于零.

在各种不同的曲面上，测地线的情况可以是种类繁多的．在图 41 上举例画着旋转双曲面上的若干条测地线.

曲面的弯曲变形　研究曲面在弯曲变形下不改变在性质的内蕴几何，密切地联系着弯曲变形本身的研究．曲面的弯曲变形的理论属于几何学中最富有内容和最困难的部分，它包含着大量的问题，虽然其中有一些问题的提法简单而又自然，但是直到今天还没有获得最终的解决.

欧拉和闵定格也研究过曲面的弯曲变形的问题，但是关于任意曲面的弯曲变形的普遍的结果是在后来才得到的.

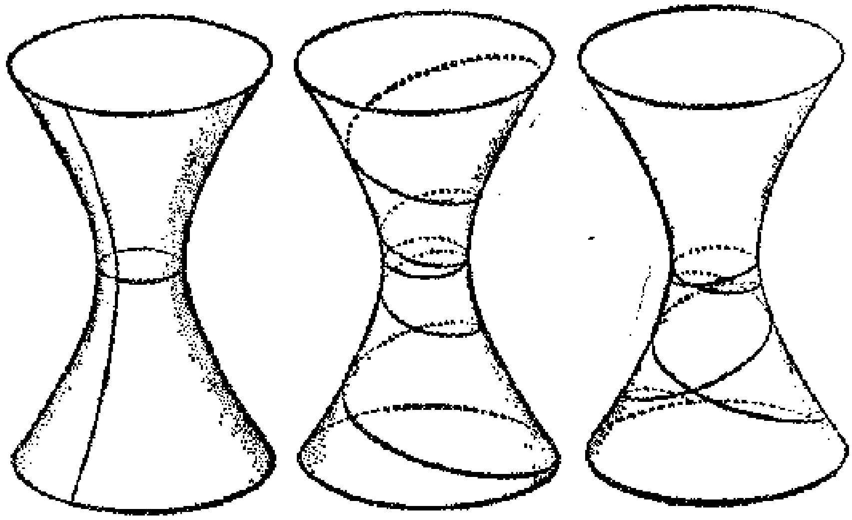


图 41

在弯曲变形的一般理论里首先发生的是曲面是否总能弯曲变形的问题，而且假如能够弯曲变形的话，那末这种弯曲变形又具有怎样的自由度？对于所谓解析曲面，即可以用能展成泰勒级数的坐标函数给定的曲面，这个问题在上一世纪末叶由法国几何学家达尔布解决了．特别地有以下的事实：如果在曲面上取定任意一条测地线而且在空间中给定任意一条有同样长度但是到处没有零曲率的(解析)曲线，那么包含已知测地线的充分狭窄的曲面长条就可以作这样的弯曲变形，使得测地线变成给定的曲线[1]．这个定理表明，曲面长条可以作完全任意的弯曲变形．然而还可以证明，如果测地线所应该变成的曲线已经给定，则曲面最多可以用两种方式作弯曲变形．(举例说，如果这曲线是平面曲线，则曲面的两个位置就正好对称于这个平面．)而如果测地线是直线，则最后这个命题不成立，因为可以举柱状曲面的任意弯曲变形作为例子来说明这一点．

我们定义弯曲变形作为曲面的这样的变形，经过这种变形，曲面上的所有曲线的长度都保留不变．这时说的是变形的最终结果；至于曲面本身在变形过程中将会变得怎样的问题则没有牵

1）把测地线变成具有零曲率的曲线的情形是例外的，因为不难肯定，对于高斯曲率为正数的曲面，这显然是不可能的．

涉到．然而假如把曲面看做由能弯曲而不能伸缩的材料做成的，则自然可以讨论在变形的每一时刻长度都保留不变(在物理学上这就相当于无伸缩性的物体)的连续的变形．这种变形叫做连续的弯曲变形．

初看起来，可以认为，任何连续变形都能用连续的过程来实现，然而这并非事实．例如可以证明，有圆槽形状的曲面(图 42)就不允许连续的弯曲变形(这同时说明了这样一个大家知道的事实，圆形的水桶要比方形的水桶来得坚固)，虽然这种曲面是可以作弯曲变形的：只要沿着它与水平面相切的圆把它割开，把其中的一片换成它对于这平面的镜面反射图形(在图 43 上，正像在图 42 上一样，为了直观起见，只画出所讨论的曲面的一半)．圆槽的圆环形状阻碍它作连续的弯曲变形从直观上看来是很明显的(而像直槽那样的曲面，相当于上面所写的那种弯曲变形是可以连续地完成的)．

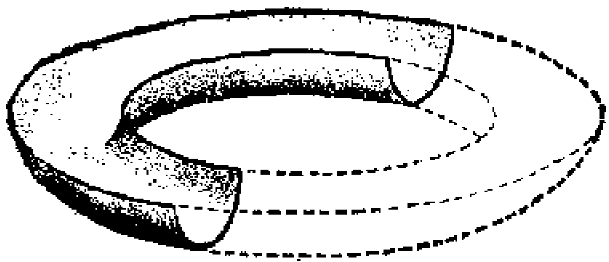


图 42

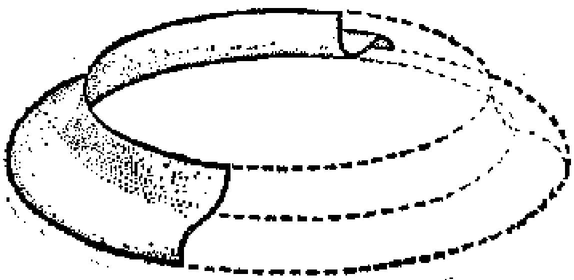


图 43

假如限于曲面的充分小的片段，那么要它作连续的弯曲变形就没有什么看得见的障碍了，而且可以意想到，曲面的微小片段的任何弯曲变形总可以用连续的方式实现，只是在必要时需要补充

一个镜面反射．这确实是对的，但是必须在曲面的所讨论片段上高斯曲率不成为零的条件下(不算它处处等于零的情形)．而如果高斯曲率在一些个别的点处变成零，则就像叶菲莫夫在 1940 年所证明了的，即使是曲面的充分小的片段，也可以不许可作保留正则性的弯曲变形．举例说，由方程 $z=x^9+\lambda x^6y^3+y^9$ (这里 λ 是超越数)给定的曲面就是这样，它的任何包含坐标原点的(即使是充分小的)片段都不许可作充分正则的连续的弯曲变形．叶菲莫夫定理是古典微分几何中的新的和有些出于意外的成果．

与弯曲变形理论的一般问题同时，在几何学中有很多地方是研究曲面的一些特殊类型的弯曲变形的．

曲面的内蕴几何与它的空间形式的联系 我们已经知道，曲面和曲面上的图形的某些直接与它的空间形式有联系的性质，即某些所谓"外在几何的"性质，由曲面的内蕴几何决定．举例说，曲面的主曲率的乘积(高斯曲率)就由它的内蕴几何决定．另一个例子：对于处在曲面上的曲线，主法线到处与曲面法线重合的必要和充分的条件是，这曲线具有确定的内蕴几何的性质，它即是测地线．

曲面的空间形式只是以一定的任意性随着曲面的这种内蕴几何而被决定．

曲面的空间形式对它的内蕴几何的依赖性可以解析地用方程来表出，在这些方程里出现刻画内蕴几何的量和刻画曲面的外在的弯曲性的量．这些公式中的一个是由高斯提出的，那就是用属于内蕴几何的量来表示高斯曲率的公式．另外两个方程是在 §1 里提到过的彼德逊-科达奇方程．

高斯-彼德逊-科达奇方程完全地包括了在曲面的内蕴几何和它在空间的弯曲性之间的联系，因此在任意曲面的内蕴几何的量和外在几何的量之间的所有可能的关系，至少是以隐蔽的形式，已经包括在这些方程之中了．

由于曲面在空间中的形式不单由它的内蕴几何决定，自然会发生这样的问题：为了完全决定曲面，还需要给定哪样一些外在几

何的量？原来，如果两个曲面有同样的内蕴几何，而且这两个曲面在对应的点和方向上的法截线的曲率在取定符号后是相等的，那么这两个曲面本身也就相等，即可以通过运动而使它们叠合.我们应该指出，彼德逊比波奈早15年发现这个定理，虽然这定理常常以后者来命名.

曲面理论的解析的工具 分析学在曲面理论里的系统的应用引导出特别适合于这个目的的解析工具的创立. 在这方面的有决定意义的一步是由高斯引入的，那就是用所谓曲纹坐标解析地给定曲面的方法. 这个方法是平面上的笛卡儿坐标的观念的自然推广，而且与曲面的内蕴几何密切有关，在内蕴几何里用方程 $z=f(x, y)$ 来给定曲面已经变得不方便了. 不方便的原因在于曲面上的点的坐标 x, y 在弯曲变形下要改变. 为了消除这种不方便，人们在曲面本身之上引进坐标，用与每个点有关的两个数 u, v 来决定这个点(这些数在经过弯曲变形后还与已知点有关). 空间的坐标 x, y, z 每每就是 u, v 的这个或那个函数了. 在曲面上给定一个点的数 u 和 v 叫做这个点的曲纹坐标. 这个概念是这样命名的：如果固定其中一个坐标(例如 v)的值，而让另一个坐标变动，那末我们就在曲面上得到一条坐标曲线. 坐标曲线在曲面上组成曲线网，就像平面上的坐标网似的. 我们注意到，在地球面上用经度和纬度决定点的位置的熟知方法并非别的，而正是在球面上引进的曲纹坐标；在这情形里坐标网由圆周——经线和纬线组成[1](图 44). 为了用曲纹坐标来给定曲面的空间位置，需要决定曲面的点依赖于 u 和 v 的情况，例如把从某个固定的起点引向曲面的点的向量 $\boldsymbol{r}$(所谓曲面的向径) 给定为 u 和 v 的函数，$\boldsymbol{r}=\boldsymbol{r}(u, v)$ (这等价于把向量 $\boldsymbol{r}$ 的分量 x, y, z 给定为 u 和 v 的函数)[2]. 为

---

1) 特别是，早在笛卡儿在平面上引入通常的坐标以前好久，地理坐标就在实际应用中出现了.

2) 高斯自然并没有利用向量的记号. 他直接把曲面上的点的三个坐标分别定义为 u 和 v 的函数. 在汉密登和格拉斯曼的工作成果中出现的向量，最先在物理学中得到广泛的应用，而且直到很晚(实际上是在二十世纪内)才成为讲述解析几何和微分几何时的传统的工具.

了给定处在已知曲面上的曲线，需要给定坐标 u 和 v 作为一个变数 t 的函数，在这以后，在这条曲线上变动的点的向径就可以用复合函数 $\boldsymbol{r}[u(t), v(t)]$ 表示.

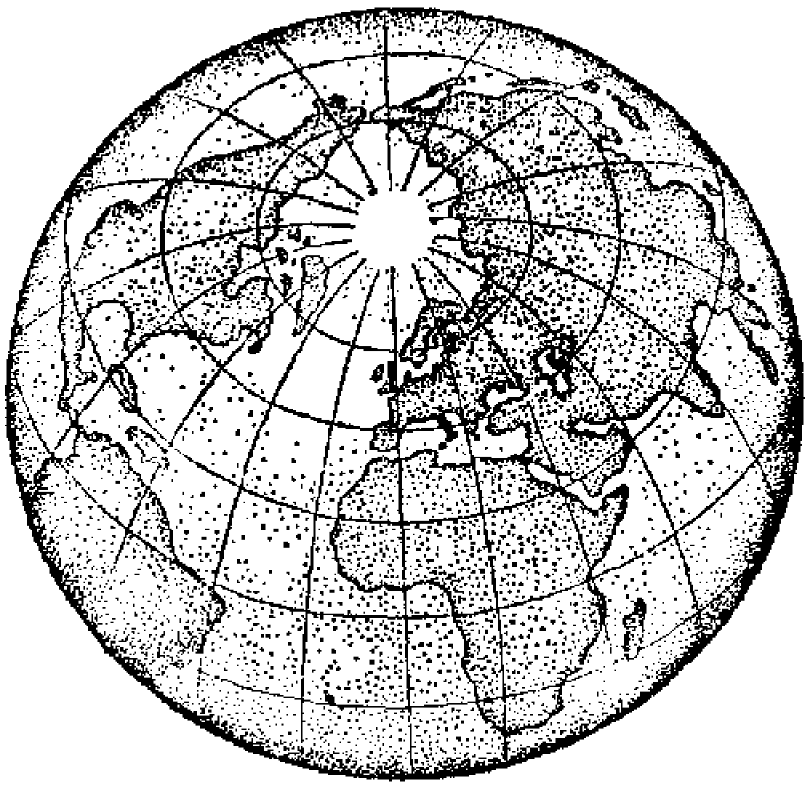


图 44

导数和微分的概念可以逐句地推广到向量函数上；这时从导数是 $\frac{\Delta\boldsymbol{r}}{\Delta t}$ 当 $t\to 0$ ($\boldsymbol{r}$ 是参数 t 的函数)时的极限的定义直接推出，曲线向径的导数是向量，它的方向沿着这条曲线的切线(图 45)．可以搬到向量上的还有普通导数的基本性质，特别是复合函数的微分法则

$$\frac{d\boldsymbol{r}[u(t), v(t)]}{dt}=\frac{\partial\boldsymbol{r}}{\partial u}\frac{du}{dt}+\frac{\partial\boldsymbol{r}}{\partial v}\frac{dv}{dt}=\boldsymbol{r}_u u_t'+\boldsymbol{r}_v v_t', \tag{7}$$

这里 $\boldsymbol{r}_u$, $\boldsymbol{r}_v$ 是向量函数 $\boldsymbol{r}(u, v)$ 的偏导数.

可以证明，曲线的长度由下列积分表示：

$$s=\int\sqrt{x'^2(t)+y'^2(t)+z'^2(t)}\,dt.$$

因此，曲线长度的微分等于

$$ds=\sqrt{x'^2(t)+y'^2(t)+z'^2(t)}\,dt.$$

但是因为 $x'(t)$, $y'(t)$, $z'(t)$ 正是向量 $\frac{d\boldsymbol{r}}{dt}=\boldsymbol{r}_t'$ 的分量，所以可以写成

$ds=|\boldsymbol{r}_t'|dt$，这里$|\boldsymbol{r}_t'|$表示向量 $\boldsymbol{r}_t'$ 的长度．对于处在曲面上的曲线，按照(7)我们得到

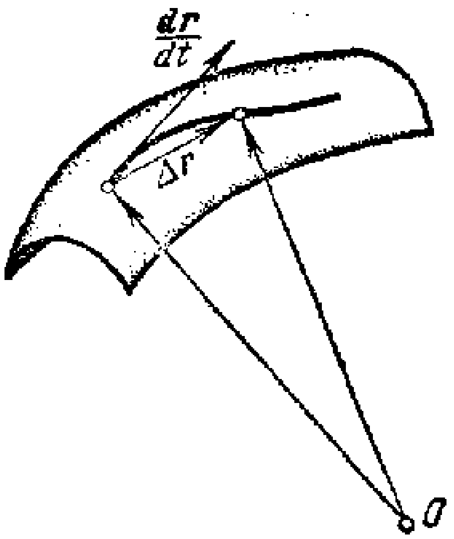


图 45

$$ds=|\boldsymbol{r}_u u_t'+\boldsymbol{r}_v v_t'|dt.$$

按照向量代数的法则计算等式右边的向量长度的平方[1)]，我们得到

$$ds^2=[\boldsymbol{r}_u^2 u_t'^2+2\boldsymbol{r}_u\boldsymbol{r}_v u_t'v_t'+\boldsymbol{r}_v^2 v_t'^2]dt^2.$$

换成 u 和 v 的微分，而且引进记号

$$\boldsymbol{r}_u^2=E(u,\ v),$$

$$\boldsymbol{r}_u\boldsymbol{r}_v=F(u,\ v),$$

$$\boldsymbol{r}_v^2=G(u,\ v),$$

我们就有

$$ds^2=E\,du^2+2F\,du\,dv+G\,dv^2.$$

我们看到，曲面上弧长微分的平方是微分 du 和 dv 的二次形式，其系数与曲面的点有关．这个二次形式叫做曲面的第一个二次形式．在曲面的每个点处给定了第一个二次形式的系数E, F, G 以后，就可以按照公式

$$s=\int_{t_1}^{t_2}\sqrt{Eu_t'^2+2Fu_t'v_t'+Gv_t'^2}\,dt$$

来计算任意曲线的长度，因此也就完全决定了曲面的内蕴几何．

作为例子，我们来说明角和面积如何用 E, F, G 来表示．设从一个点引出两条曲线，其中一条由方程 $u=u_1(t)$，$v=v_1(t)$给定，而另一条由方程 $u=u_2(t)$，$v=v_2(t)$给定．于是这两条曲线的切向量就是向量

$$\boldsymbol{r}_1=\boldsymbol{r}_u\frac{du_1}{dt}+\boldsymbol{r}_v\frac{dv_1}{dt},$$

$$\boldsymbol{r}_2=\boldsymbol{r}_u\frac{du_2}{dt}+\boldsymbol{r}_v\frac{dv_2}{dt}.$$

这两个向量之间的角的余弦等于它们的数量乘积 $\boldsymbol{r}_1\boldsymbol{r}_2$ 被它们的长度的乘积 r_1r_2 除，

$$\cos\alpha=\frac{\boldsymbol{r}_1\boldsymbol{r}_2}{r_1r_2}$$

$$=\frac{\boldsymbol{r}_u^2\frac{du_1}{dt}\frac{du_2}{dt}+\boldsymbol{r}_u\boldsymbol{r}_v\left(\frac{du_1}{dt}\frac{dv_2}{dt}+\frac{du_2}{dt}\frac{dv_1}{dt}\right)+\boldsymbol{r}_v^2\frac{dv_1}{dt}\frac{dv_2}{dt}}{r_1r_2}$$

1) 向量长度的平方等于向量与其自己的数量乘积，而对于数量乘积，我们知道[参看第三章(第一卷)§9]，普通去括弧的法则成立．

回想到 $\boldsymbol{r}_u^2=E$, $\boldsymbol{r}_u\boldsymbol{r}_v=F$, $\boldsymbol{r}_v^2=G$ 我们得到

$$\cos\alpha=\frac{E\frac{du_1}{dt}\frac{du_2}{dt}+F\left(\frac{du_1}{dt}\frac{dv_2}{dt}+\frac{du_2}{dt}\frac{dv_1}{dt}\right)+G\frac{dv_1}{dt}\frac{dv_2}{dt}}{\sqrt{E\left(\frac{du_1}{dt}\right)^2+2F\frac{du_1}{dt}\frac{dv_1}{dt}+G\left(\frac{dv_1}{dt}\right)^2}}$$

$$\cdot\frac{1}{\sqrt{E\left(\frac{du_2}{dt}\right)^2+2F\frac{du_2}{dt}\frac{dv_2}{dt}+G\left(\frac{dv_2}{dt}\right)^2}}.$$

为了得到面积的公式，我们来讨论曲边四边形，设它的各边是曲线 $u=u_0$, $v=v_0$, $u=u_0+\Delta u$, $v=v_0+\Delta v$. 我们近似地把它换成切平面上的一个平行四边形，这个平行四边形是以切于坐标曲线的向量 $\boldsymbol{r}_u\Delta u$ 和 $\boldsymbol{r}_v\Delta v$ 为边的(图 46)．这个平行四边形的面积是 $\Delta s=|\boldsymbol{r}u||\boldsymbol{r}v|\Delta u\Delta v\sin\varphi$, 这里 φ 是在 $\boldsymbol{r}_u$

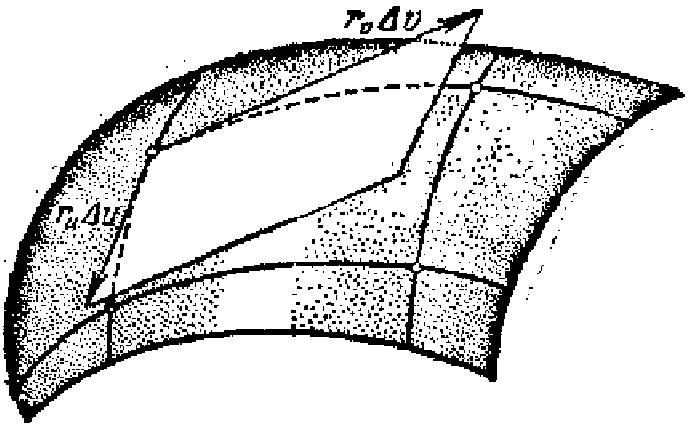


图 46

和 $\boldsymbol{r}_v$ 之间的角．因为 $\sin\varphi=\sqrt{1-\cos^2\varphi}$, 所以

$$\Delta s=|\boldsymbol{r}_u||\boldsymbol{r}_v|\Delta u\Delta v\sqrt{1-\cos^2\varphi}=\sqrt{\boldsymbol{r}_u^2\boldsymbol{r}_v^2-|\boldsymbol{r}_u|^2|\boldsymbol{r}_v|^2\cos^2\varphi}\,\Delta u\,\Delta v.$$

回想到， $\boldsymbol{r}_u^2=E$, $\boldsymbol{r}_v^2=G$, $|\boldsymbol{r}_u|\cdot|\boldsymbol{r}_v|\cos\varphi=\boldsymbol{r}_u\boldsymbol{r}_v=F$, 我们得到 $\Delta s=\sqrt{EG-F^2}\Delta u\Delta v$. 把平行四边形的面积加起来，而且在 Δu, $\Delta v\to 0$ 时过渡到极限，我们就得到面积的公式

$$S=\iint_D\sqrt{EG-F^2}\,du\,dv.$$

这里积分是对于与已知曲面片段对应的变数 u, v 的变化区域 D 进行的．

因此，曲纹坐标在研究曲面的内蕴几何时是非常适用的．

原来曲面在空间中的弯曲性也可以用微分 du 和 dv 的一个二次形式来刻画．事实上，假如 $\boldsymbol{n}$ 是曲面在点 M 处的单位法向量，而 $\Delta\boldsymbol{r}$ 是当从点挪动一下时曲面的向径的改变量，那么曲面与切平面的差异 h(图 47) 就等于 $\boldsymbol{n}\Delta\boldsymbol{r}$．按泰勒公式展开 $\Delta\boldsymbol{r}$, 我们得到

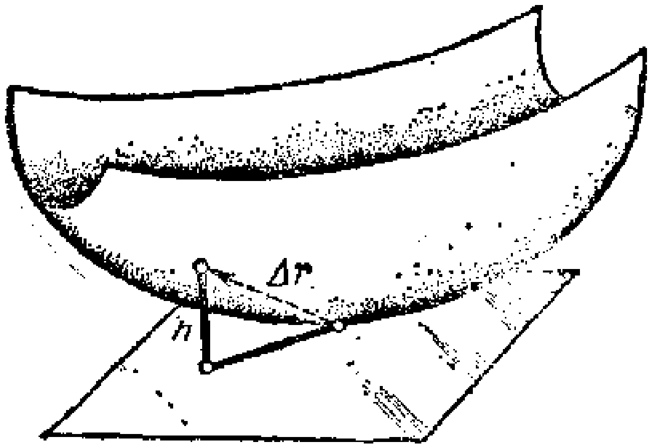


图 47

$$h = \boldsymbol{n}d\boldsymbol{r} + \frac{1}{2}\boldsymbol{n}d^2\boldsymbol{r} + \varepsilon(du^2 + dv^2),$$

这里当$\sqrt{du^2+dv^2}\to 0$时有 $\varepsilon\to 0$. 因向量 $d\boldsymbol{r}$ 处在切平面上，所以 $\boldsymbol{n}d\boldsymbol{r}=0$; 最后一项 $\varepsilon(du^2+dv^2)$与微分 du 和 dv 的平方相比是微小的. 留下的是主要项 $\frac{1}{2}\boldsymbol{n}d^2\boldsymbol{r}$. 两倍于$h$ 的主要部分的量 $\boldsymbol{n}d^2\boldsymbol{r}$ 是 du 和 dv 的二次形式

$$\boldsymbol{n}\,d^2\boldsymbol{r} = \boldsymbol{n}\boldsymbol{r}_{uu}\,du^2 + 2\boldsymbol{n}\boldsymbol{r}_{uv}\,du\,dv + \boldsymbol{n}\boldsymbol{r}_{vv}\,dv^2.$$

这个二次形式就表明曲面与切平面的差异的特征. 它叫做曲面的第二个二次形式. 它的依赖于 u 和 v 的系数常常记做

$$\boldsymbol{n}\boldsymbol{r}_{uu}=L,\ \boldsymbol{n}\boldsymbol{r}_{uv}=M,\ \boldsymbol{n}\boldsymbol{r}_{vv}=N.$$

知道了第二个二次形式，我们就可以计算曲面上的任意曲线的曲率. 事实上，应用公式 $k=\lim\limits_{l\to 0}\frac{2h}{l^2}$, 我们得到，方向与微分比值 $du:dv$ 对应的法截线的曲率等于

$$k_n = \frac{\boldsymbol{n}d^2\boldsymbol{r}}{ds^2} = \frac{L\,du^2+2M\,du\,dv+N\,dv^2}{E\,du^2+2F\,du\,dv+G\,dv^2}.$$

如果所讨论的曲线不是法截线，那么按照梅涅定理，只要用向着同一个方向的法截线的曲率除以曲线的主法线和曲面法线之间的角的余弦就成.

第二个二次形式的引入给出了研究曲面的弯曲性的解析方法. 特别地，就可以纯解析地获得欧拉定理和梅涅定理，获得高斯曲率和平均曲率的表示式等等.

前面说过的彼德逊定理表明，同时取定的两个二次形式决定一个曲面只差它在空间中的位置，因此，曲面的任意性质的研究从解析方面看就化成研究这两个二次形式. 最后我们要指出，两个二次形式的系数并不是互不相关的; 前面提到过的在曲面的内蕴几何和它的空间形式之间的联系，解析地就

以第一个和第二个二次形式的系数之间的三个关系(高斯-科达奇方程)表出.

§5. 曲线和曲面理论中的新方向

曲线族和曲面族 虽然曲线和曲面理论的原理在很大程度上在上一世纪中叶已经完成了,但是这理论过去继续有发展,今天还在继续发展着. 同时在这理论中逐渐增加了一系列新的方向,扩大了在现代微分几何中所研究的图形和其性质的范围. 不过这些方面中的一个早在微分几何的萌芽时期就已经出现了. 我们指的是"族"的理论,即曲线和曲面的连续集合的理论. 但是因为直到曲线和曲面理论的原理已经完全形成时,"族"的理论才开始了深入的发展,在这意义下可以认为它是新的方向.

图形的连续集合一般叫做 **n 参数的族**,假如这集合的每个图形由 n 个参数的值给定,并且假定所有刻画一个图形的量(关于图形的位置、形状等等)至少是连续地随着这些参数而改变的. 从这个普遍定义的观点看来,可以认为曲线是单参数的点族,而曲面则是二参数的点族. 因为平面上的圆由三个参数给定:中心的两个坐标和半径,所以平面上的所有圆周的集合就是三参数的曲线族.

曲线族或曲面族的理论中的最简单问题是求出所谓族的包络的问题. 假如一个曲面在其每个点处与一个曲面族中的一个曲面相切,而且它以这种方式与曲面族中的每一个曲面相切,这曲面就叫做这曲面族的包络. 譬如说,中心的已知直线上的等半径的球面族的包络是圆柱面(图 48),而以已知平面的所有点为中心的这种球面族的包络则由两个平行平面组成. 同样地可以定义曲线族的包络. 在图 49 上画着以不同的角度从喷泉中喷出的水流;在同一个平面上的这种水流形成一个曲线族,这些曲线可以近似地认为是抛物线. 在照片上可以清楚地看出它们的包络,它好象是水冠的周界. 当然,并不是每一个曲线族或曲面族都有包络(举例说,平行的直线族就没有包络). 求任意族的包络的简单而普遍的

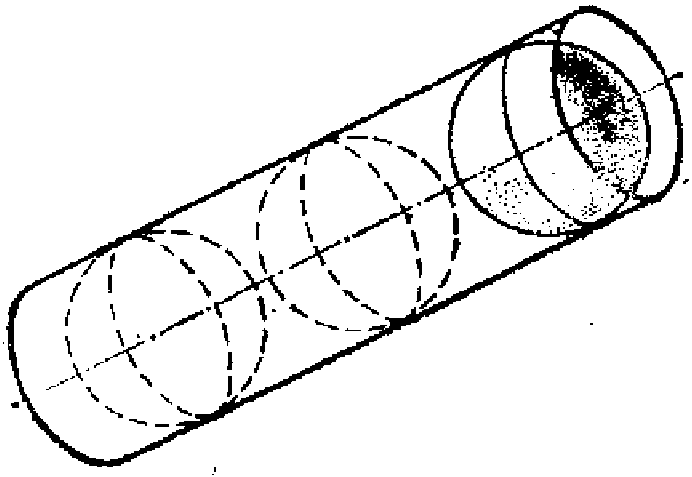


图 48

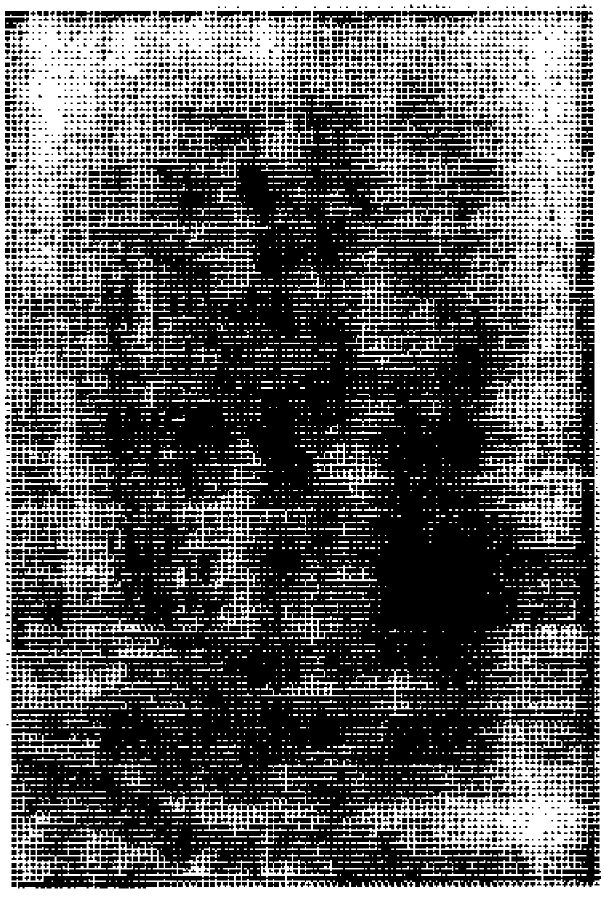


图 49 (a)

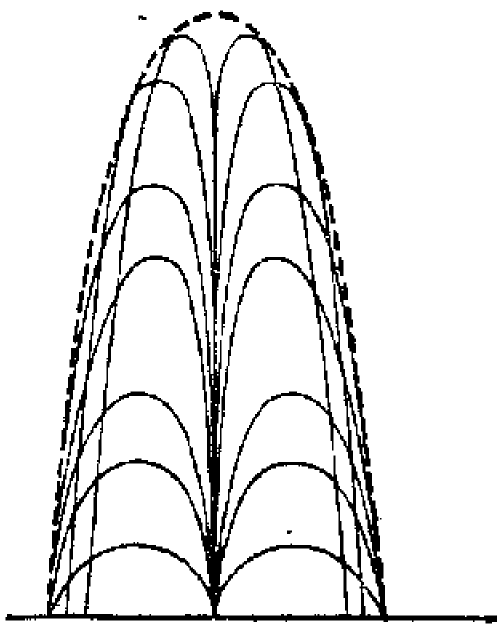


图 49 (b)

方法是有的；对于平面上的曲线族，这种方法还是莱布尼兹所给出的哩.

每一条曲线显然是它自己的切线的包络. 同理，每一个曲面是它自己的切平面的包络. (这恰好是用给定曲线的切线族来给

定曲线或者用给定曲面的切平面族来给定曲面的新方法的根据。在很多问题中，这种给定曲线或曲面的方法是很方便的.)

一般地说，在曲面的不同的点处，切平面是不同的，因此曲面的切平面族通常是二参数的. 但是在某些情形下，例如在柱面的情形，切平面族是单参数的. 可以证明下面的重要定理成立. 有单参数的切平面族的曲面是而且只是可以摊开到平面上的曲面，即它的任何充分小的片段都可以弯曲成平面的片段. 这就是我们在 § 4 中提到过的可展曲面. 每个这种样子的解析曲面都由直线段组成，而且或者是平面(平行直线组成)，或者是锥面(聚集于一个点的直线组成)，或者由某一条空间曲线的切线组成.

包络理论常常在工程问题(例如在传动理论)里利用. 我们来讨论两个齿轮 *A* 和 *Б*. 它们的相对运动可以这样设想：假定齿轮 *A* 是固定的，而齿轮 *Б* 则沿着它滚动(图 50). 于是齿轮 *Б* 的周界在取不同的位置时就在齿轮 *A* 的平面上形成一个曲线族，而齿轮 *A* 的周界应该总与它们相切，即是这族的包络. 当然，对于传动来说这还是不够的，还应该考虑从一对齿到另一对齿的衔接，但是上面指出的条件毕竟是主要的，齿轮的形状总应该满足这个条件.

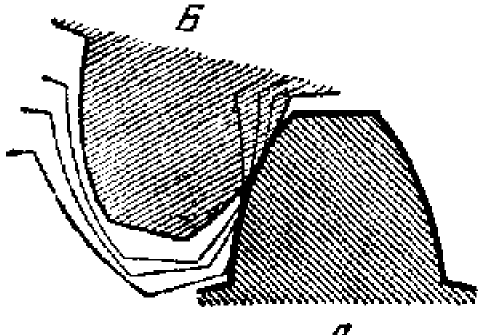


图 50

我们曾经说过，包络的问题只不过是曲线族和曲面族理论中最简单的和早已解决了的问题. 族的理论就其可能包含的问题来说不见得会比例如曲面理论更窄些. 探讨得特别多的是“线汇”理论，即各种二参数的曲线族(特别是直线族，所谓直线汇)的理论. 在这理论中应用的方法实质上与曲面理论中所用的方法相同.

直线汇的理论早在孟日的著作“论沟和堤”中已经开始，这著作的名称表明，孟日的研究是在联系到实际问题时产生的；谈到的是从沟到堤上的最便利的运土法.

在上一世纪中叶才开始的线汇理论的系统发展，在很大程度

上是与几何光学有关的；因为在均匀介质中的光线的集合常常是直线汇．

非正则曲面和"整体"几何 曲线和曲面(以及曲线族和曲面族)的理论是在上世纪末叶形成的，通常把它叫做古典微分几何；作为它的标志的特别是以下的特征．

首先，它只讨论"充分平滑的"即所谓正规的曲线和曲面，也就是说它们是由具有足够多个导数的函数所给定的．因而像具有尖端和棱角的曲面，例如多面体表面和锥面，不是完全排斥到讨论之外，就是只讨论曲面上平滑的片段了．

其次，古典微分几何专门研究曲线和曲面的充分小的片段的性质("局部"的几何学)，而完全不关心整个闭曲面有什么性质这种问题(这种问题已经属于所谓"整体"的几何学了)．

揭露"局部"几何和"整体"几何的差别的典型例子可以由曲面的弯曲变形来给出．举例说，早在1838年，闵定格就证明了球面的充分小的片段可以作弯曲变形("局部"的定理)．同时它还提出了整个球面不能作弯曲变形的假设．这后一个定理直到1899年才由别的数学家证明了．恰好，对于由可以弯曲但是确实不能伸展的材料做成的球面，我们很容易用实验来揭示它的不可弯曲性．举例说，虽然乒乓球是用很容易弯曲的材料做成的，但是它却是很坚固的．再有一个例子是我们在§4里已经说到过的普通的洋铁水桶；由于水桶具有圆形的边缘，它整个地说是坚固的；然而同一个曲面的个别片段却是可以弯曲的．如同我们所看到的，在曲面的"局部"性质和"整体"性质之间可以发现实质上的差别．

我们在§4里获得过初步知识的测地线理论，也给我们一个特征性的例子．测地线在"局部"上，即在小片段上，乃是最短线，但是它"整体"地可以完全不是最短线，而且可以像在球面上大圆的情形一样是封闭的．

读者不难注意到写在§4中的关于测地线的定理基本上是"局部"的定理．而关于测地线的整个伸展情况的问题则是属于整体几何的．举例说，大家知道，在正则曲面上的一个很小的邻域

内，只有唯一的测地线连接两个充分接近的点．而假如从测地线的伸展范围上离开这两个点非常远的地方来看测地线，则按照莫尔斯定理，在闭曲面上的任意一对点都可以用无穷多条测地线把它们连结起来．举例说，圆柱侧面上的两个点 A 和 B 就可以用完全不同的测地线来连结它们：只要在从 A 到 B 的路径上取绕圆柱次数不同的各种螺旋线．属于“整体”几何的还有由列斯台尔尼克和史尼雷尔曼所证明的关于闭测地线的庞卡来定理，这定理将会在第十八章（第三卷）里谈到．

证明这些定理，就像证明“整体”几何中很多其他定理一样，显得是难以用古典微分几何的普通方法来做到的，因而需要创造新的方法．

正如“整体”几何的问题不可避免地会引起数学家们的注意，在科学中也不可能坚持单是正则曲面这个限制，因为我们经常遇见非正则的、不能连续地弯曲变形的曲面，例如立方体、圆锥、有尖的边缘的凸透镜等等的表面．

此外，很多解析曲面在自然扩展下必然会出现“奇异性”——违反了正则性，例如出现棱角和尖端．譬如圆锥面的片段在扩展后就要引向顶点——尖端，在那儿是违反曲面的平滑性的．

最后提到的事实只不过是下列重要定理的特别情形．不是柱面的任何可展曲面在自然的扩展下总要遇到棱角（在锥面情形是尖端），在遇到棱角时这曲面已经不能在保持正则性的条件下再扩展了．

因此，在曲面的“整体”情况和它的奇异性之间有很深的联系．这正是为什么解决“整体”问题和研究具有“奇异性”（棱角，尖端，曲率的间断处等等）的曲面应该同时进行的原因之一．

同样的新方向也在分析学中发生．例如在第五章 §7 里说过的微分方程的定性理论，就研究微分方程在其整个定义区域里的解的性质，即所谓“整体”的性质．因而它特别注意“奇异性”，注意正则性的破坏，注意方程的奇异点．此外，现代的分析学所研究的对象包括了非正则的函数，而它正是古典分析学所不研究的（参看

第三卷第十五章)；这就使几何学在研究更一般的曲面时有了新的工具．最后，在通常处理具有某种极值性质的曲线或曲面的变分法里，有时达到极值的极限曲线会显得不是正则曲线．在这种问题的提法中所必须的曲线类或曲面类的封闭性，也要求推广所讨论的曲线或曲面的范围——首先至少得包括最简单的非正则曲线或曲面．总之，几何学中的新方向的产生并不是孤立的，而是与数学的整个发展密切地有联系的．

大约在50年前才开始解决“整体”问题和不仅研究正则曲面的转变．有很多数学家参加了这个新方向的研究．赫尔曼·闵可夫斯基(1864—1909)作出了第一个重要的贡献，建立了几何学中宽广的一章——凸体理论的初步．顺便说一下，促使闵可夫斯基进行这些研究的问题之一是关于正方格子的问题，这问题是与数论和几何结晶学有密切联系的．

一个立体叫做凸的，假如通过它的表面的每个点，都可引不会截割这个立体的平面，即这个立体可以用它表面的任何点支持在平面上(图51)．凸体由它的表面决定，而且在很大程度上当我们说到凸体理论和闭凸曲面理论时是没有差别的．关于凸体的普遍定理的真实性，通常并不需对它的表面的平滑性或“正则性”作任何补充的假定．即是说这些定理通常属于整个的凸体或整个的凸曲面．因此，凸体和凸曲面的理论在其基本方面就打破了古典微分几何的限制．然而凸体和凸曲面的理论还稍稍与古典微分几何有些联系；这两种理论的统一是很晚才发生的．

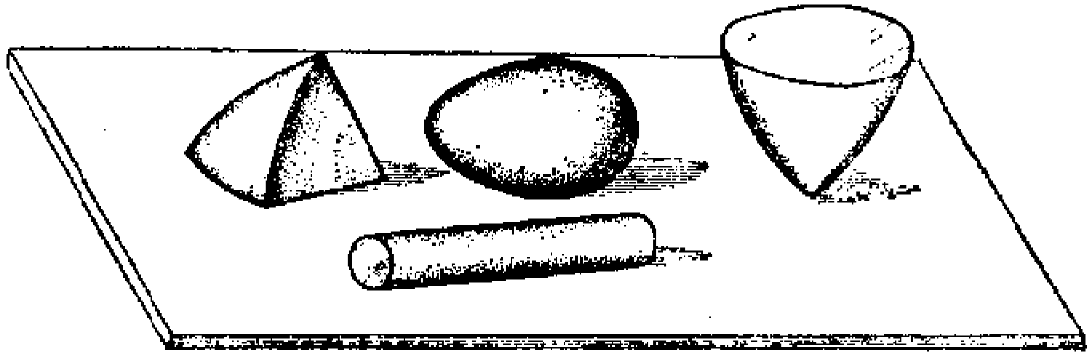


图 51

从1940年起，A. Д. 亚历山大洛夫发展了一般曲线和曲面的理论. 这理论不仅包括了古典微分几何的正则曲面，而且也包括了非平滑的曲面：多面体表面，任意凸曲面和其他曲面. 虽然问题的提法和这理论的结论有很大的普遍性，但是作为这理论的基础的首先是直观的几何概念和几何方法，当然它确实还要利用现代分析学的概念和方法. 这理论的基本方法之一是用多面形（多面体表面）逼近一般的曲面. 这个方法的最简单情形是每一个中学生都知道的，譬如把圆柱的侧面积当做角柱面积的极限来计算就是如此. 在一系列的情形里这方法给出强有力的结果，这些结果或者是用别的方法所不能得到的，或者是在引向极限的解析过程中需要很大的努力和引起复杂的思考. 这个方法是这样的，先对多面形解决了问题，然后用过渡到极限的方法把结果搬到一般的曲面上.

关于从已知的摊开图形粘合成一个凸多面形的条件的定理，乃是一般凸曲面理论的原理之一. 这个定理从叙述上看完全是初等的，然而却有非初等的证明而且引出关于一般凸曲面的极为深入的推论. 读者当然知道如何粘合成立方形的盒子，例如，由十字形的纸样来粘合成立方形(图52). 这个用纸片粘合多面形的方法可发展成为"摊开和粘合"的一般方法，这方法在曲面理论的多种问题中给出强有力的结果而适宜于实际应用.

这理论中的一些深入的结果是由波果列洛夫获得的. 特别地，他证明了任意闭凸曲面作为一个整体是不能在保留它的凸性的条件下作弯曲变形的. 在1949年获得的这个结果结束了很多著名的数学家的工作，这些数学家在50年的长时期中企图证明这个定理，但是只在这样或那样的补充条件下才得到它的证明. 波果列洛夫的结果在结合了"摊开和粘合"的方法以后，不仅给出了这个问题的完全的解，而且几乎是最终地阐明了关于闭的和不闭的凸曲面的可弯曲性和不可弯曲性的全部问题. 他还确立了新的理论与"古典的"理论的密切联系.

这样一来，终于建立了这样的曲面理论，它不仅包括了古典

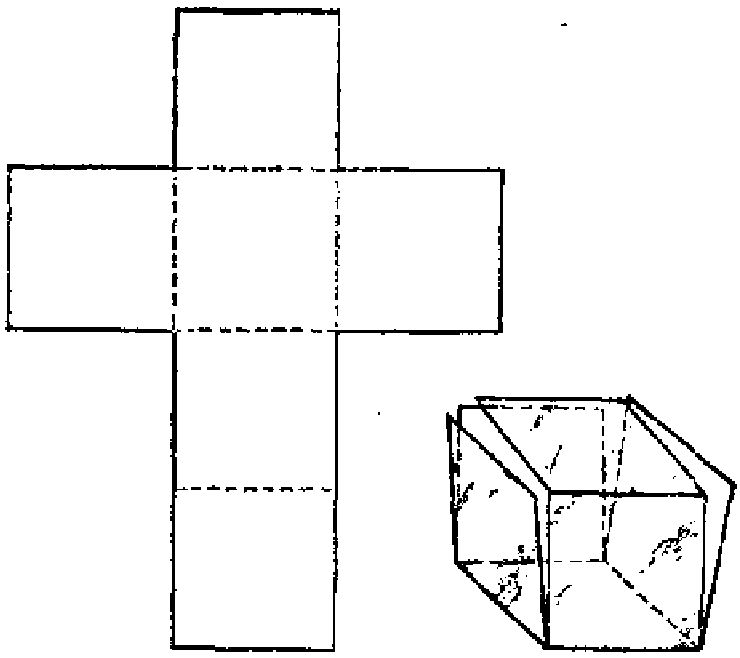


图 52

的理论，而且包括了多面形、任意凸曲面和极为一般的非凸曲面的理论。遗憾的是我们不能在这里详细地说一下这个理论中的结论和尚未解决的问题，虽然这一点是可以做到的，因为它们很多都是极为直观的，而且尽管它们的严密证明是困难的，为了理解问题的提法却并不需要特别的知识。

前面在谈到曲面的弯曲变形的§4里，我们曾处理过正则的（连续地弯曲的）曲面在保留其正则性的条件下的弯曲变形．相反地，在刚才提到的波果列洛夫定理里，并不要求原来的和弯曲成的曲面的正则性，但却添上了两个曲面都是凸曲面的要求。

明显地，假如允许得到的曲面不必是凸曲面和可出现绉折，则球面的弯曲变形就成为可能的了．只要从球面割一块而把它从原来的位置变成翻转过来的形状，即像把球面的片段压向里面去一样，我们就得到这种弯曲变形的一个例子．不久以前由美国数学家涅什和荷兰数学家凯伯所获得的结果是特别出于意外的．他们证明了，假如只保留曲面的平滑性而允许在曲面的曲率中出现某种跳跃（即对于已知函数撤去了二阶导数的连续性、有界性甚至存在性的任何要求），则曲面作为一个整体就可能以很大程度的任意

性作弯曲变形了．特别地，球面片段就可以弯曲成某种小块，这种小块是具有很小的波浪状绉纹的平滑曲面．通过关于这种变形的某种认识，就可清楚地想像由柔软物质（例如布）做成的球形口袋是能够极为任意地弄绉的．至于赛璐珞的小球则是另一回事了．它的弹性表面不但不允许伸展，而且也不允许突然的弯曲，因而这种小球就显得是非常坚固的了．

各种变换群的微分几何 本世纪的开端从古典微分几何发掘出一系列新方向，它们是统一在一个观点之下的．这指的是专门研究曲线、曲面和曲线族、曲面族的这样一些性质，它们是在这种或那种形式的变换下不变的性质．古典微分几何研究在运动下不变的性质，但是自然可以考虑其他的几何变换．例如，在空间范围里把直线还变成直线的变换叫做射影变换．很久以前已经产生了所谓射影几何，它研究在任意射影变换下保留不变的图形的性质．射影几何就其所研究的问题说是与通常的初等几何和解析几何相类的，直到本世纪开始才在很多数学家的工作中开始发展“射影微分几何”，即类似于古典微分几何的曲线、曲面和曲线族、曲面族的理论，只是专门研究它们在任意射影变换下保留着的性质．这方向的创始工作是美国数学家维尔琴斯基的，以及意大利数学家富比尼和捷克数学家切赫的工作．

同样地还产生了“仿射微分几何”，它研究曲线、曲面和曲线族、曲面族在仿射变换下保留的性质，所谓仿射变换是不但把直线变成直线，而且不违反直线的平行性的变换．德国数学家伯拉什略和他的学生们把几何学的这一部分发展成宽广的理论．还可以提到“保角几何”，在这种几何学中研究图形在不改变任意曲线之间的角的变换下保留的性质．

一般地说，可能有的“几何”是种类繁多的，因为其中每一个在一定的意义下都可以取任意的变换群作为基础，而且研究图形在这个群中的变换下保留的性质．定义各种“几何”的这种原则我们还将在第十七章（第三卷）里谈到．

苏联数学家们（菲尼可夫，拉普捷夫和其他人）也有成就地发

展了和发展着微分几何的新方向．但是在我们这一篇概述中无法描述造成微分几何各个方面成就的现代研究的全貌．

文　　献

Д. Гильберти, С. Кон-Фоссен, Наглядная геометрия., Гостехиздат, 1951.

概括大量的几何结果，文字通俗，内容丰富．有一章专门讲述微分几何．

Л. А. Люстерник, Кратчайшие линии (вариационные задачи), Популярные лекции по математике, вып. 19. Гостехиздат, 1955.

关于测地线的书，以大量的直观材料为依据的．

大 学 教 科 书

П. К. 拉舍夫斯基，微分几何教程，高等教育出版社 1955 年版。

С. П. 菲尼可夫，微分几何教程，高等教育出版社 1954 年版。

专　门　著　作

А. Д. Александров, Внутренняя геометрия выпуклых поверхностей, Гостехиздат, 1948.

讲述凸曲面理论中的新结果的专门著作。

В. Ф. Каган, Основы теории поверхностей в тензорном изложении, Гостехиздат, ч. I, 1947; ч II; 1948.

关于曲面理论的篇幅很大的著作，在其中用统一的方法叙述了大量古典的和近代的成果。

袁光明 译

孙以丰 校

第八章 变 分 法

§1. 绪 论

变分问题举例 为了弄清楚在变分法[1]中所研究的问题的范围，我们先来考虑一些个别的问题.

1. 最快滑行曲线 最速降线或最快滑行曲线问题是历史上变分法开始发展的第一个问题.

在所有连接 M_1 及 M_2 两点的曲线之中，求一曲线，使当质点在重力影响之下，没有初速地沿此曲线从 M_1 运动达到点 M_2 所需时间为最短.

为了求这个问题的解，我们应该考虑所有可能的连接 M_1 与 M_2 的曲线. 如果任取一条确定的曲线 l，它就给出了质点沿着它滑下所需时间的确定的值 T，时间 T 依赖于曲线 l 的选择，而在所有的，连接 M_1 与 M_2 的曲线之中要选出给出 T 的最小值的那一条曲线.

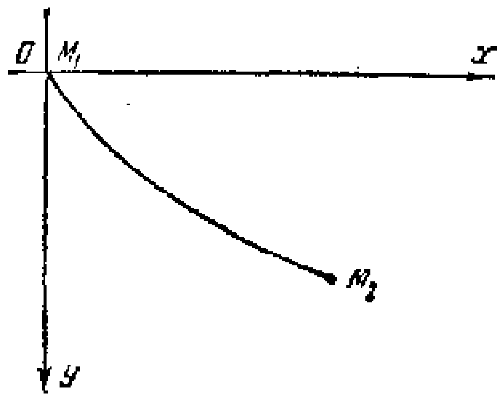


图 1

最速降线问题可以用另一形式给出.

经过点 M_1 及 M_2 作一垂直平面，显然最快滑行曲线应该在这平面内，所以求最快滑行曲线时我们就只需要限于这个平面内的曲线里来寻找了. 把点 M_1 取作坐标原点，轴 Ox 指向水平而轴 Oy 垂直向下(见图 1). 点 M_1 的坐标就是 $(0, 0)$；把点 M_2 的坐标叫做 (x_2, y_2). 任取一条可由方程

1) "变分法"这个名称的来源将在下面阐明.

$$y=f(x),\quad 0\leqslant x\leqslant x_2. \tag{1}$$

给出的曲线，其中 f 是连续可微函数. 因为曲线经过 M_1 与 M_2，函数 f 在区间 $[0, x_2]$ 的两端应满足条件

$$f(0)=0,\quad f(x_2)=y_2. \tag{2}$$

如果在曲线上取任意一点 $M(x, y)$，则质点在曲线上的这个位置上运动的速度 v 与这点的 y 坐标具有物理学中熟知的关系式

$$\frac{1}{2}v^2=gy,$$

或

$$v=\sqrt{2gy}.$$

质点走过曲线的弧长元素 ds 所需要的时间为

$$\frac{ds}{v}=\frac{\sqrt{1+y'^2}}{\sqrt{2gy}}\,dx,$$

因此质点沿曲线从 M_1 滑行到 M_2 的整个时间为

$$T=\frac{1}{\sqrt{2g}}\int_0^{x_2}\frac{\sqrt{1+y'^2}}{\sqrt{y}}\,dx. \tag{3}$$

求最速降线就和解下列极小值问题等价：在满足条件(2)的所有的函数(1)之中，求相应于积分(3)的最小值的那个函数.

2. 面积最小的旋转曲面 在连接平面上的两个点的曲线之中，求把它绕 Ox 轴旋转所得曲面面积为最小的那条曲线.

设给定的点为 $M_1(x_1, y_1)$ 及 $M_2(x_2, y_2)$，并考虑任一可由方程

$$y=f(x) \tag{4}$$

给出的曲线.

如果曲线经过 M_1 及 M_2，则函数 f 将满足条件

$$f(x_1)=y_1,\quad f(x_2)=y_2. \tag{5}$$

绕 Ox 轴旋转后，这曲线描成一个曲面，其面积的数值等于积分

$$S=2\pi\int_{x_1}^{x_2}y\sqrt{1+y'^2}\,dx. \tag{6}$$

这积分的值依赖于曲线也就是依赖于函数 $y=f(x)$ 的选择. 在所有满足条件(5)的函数(4)之中，我们要找出使积分(6)取最小值的

那个函数.

3. 变形了的薄膜的平衡 所谓薄膜就是有弹性的曲面，它在静止时是平的而只在张力作用时自由弯曲和作功. 我们认为，变形了的薄膜的位能与其表面面积的增量成比例.

设在静止状态，薄膜占有平面 Oxy 的区域 B (图 2). 把薄膜的边界 l 在垂直于 Oxy 平面的方向作变形，并用 $\varphi(M)$ 表示边上点 M 的位移. 同时薄膜的中间部分也就有了变形. 试求薄膜边界具有给定的变形时薄膜的平衡位置.

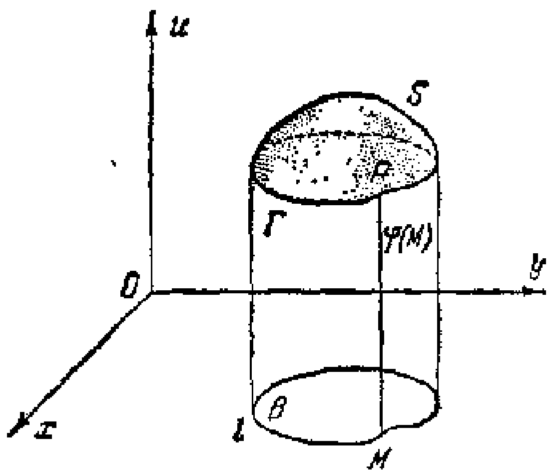


图 2

具有高度的准确性可以认为在上述变形下，所有薄膜上的点都只垂直于 Oxy 平面地移动. 以 $u(x, y)$ 记点 (x, y) 的位移. 在变形状态下，薄膜的面积为[1)]

$$\iint_B (1+u_x^2+u_y^2)^{\frac{1}{2}} dx\,dy.$$

如果薄膜元素的变形很小，以致 u_x 和 u_y 的高次幂和它们的低次幂相比较时可以略去不计，上述面积的表达式可以换成更简单的

$$\iint_B \left[1+\frac{1}{2}(u_x^2+u_y^2)\right] dx\,dy.$$

薄膜面积的变化为

$$\frac{1}{2}\iint_B [u_x^2+u_y^2]\,dx\,dy;$$

变形的位能就具有值

$$\frac{\mu}{2}\iint_B [u_x^2+u_y^2]\,dx\,dy, \tag{7}$$

1) 此处以及本章各处，偏导数都只用在函数记号的下方写出的下标表示，这个下标指明我们是对哪个自变量来求偏导数的.

式中μ是依赖于薄膜的弹性性质的常数.

由于薄膜边界上的点的位移是给定的，函数$u(x, y)$在区域B的边界l上要满足条件

$$u|_l=\varphi(M). \tag{8}$$

在平衡位置时，变形的位能具有最小的可能的值，因此确定薄膜的点的偏差的函数$u(x, y)$，应该从下列极小问题来求：在所有在区域B上连续可微和满足边界条件(8)的函数$u(x, y)$之中，求使积分(7)取最小值的函数.

泛函的极值与变分法 上面举的例子使我们有可能对在变分法中所考虑的问题的范围有所了解. 但是为了确切地来确定变分法在数学中的地位，我们应该来熟悉一些新的概念. 注意数学分析的基本概念之一是函数概念. 在最简单的情形下函数关系可以这样表示. 假定M是任何一个实数集合. 假定集合M中的每一个数x对应于某一个数y，就说在集合M上确定了函数$y=f(x)$. 集合M往往叫做函数的定义区域.

泛函的概念是函数概念的直接而又自然的推广，并且泛函的概念还包含函数概念作为它的特殊情形.

假定M是一个由任何对象组成的集合. 这些对象的本性现在对我们来说是没有关系的. 这些对象，例如，可以是空间的点、线、函数、曲面及力学体系的状态或者甚至可以是力学体系的运动，等等. 为简便计，在以后我们把它们叫做集合M的元素，而以字母x记之.

如果M的每一个元素x对应于一数y，就说在集合M上确定了泛函$y=F(x)$.

如果集合M是数x的集合，那么泛函$y=F(x)$就是具有一个自变量的函数. 当M是数偶(x_1, x_2)的集合或平面上的点的集合时，泛函就是含两个自变量的函数$y=F(x_1, x_2)$，等等.

对于泛函$y=F(x)$我们提出下列问题：

在M的所有的元素x之中，求使泛函$y=F(x)$取极小值的元素.

同样可以陈述这个泛函的极大值问题.

注意,如果把泛函 $F(x)$改号,而考虑泛函 $-F(x)$,那么 $F(x)$ 的极大值(极小值)就变成了 $-F(x)$ 的极小值(极大值). 所以把极大值和极小值分别研究是没有意义的, 因为它们的理论是非常相象的. 以后我们主要只谈泛函的极小值.

在最快滑行曲线的问题中, 我们要去找它的极小值的泛函是积分(3),也就是质点沿曲线滑行的时间. 这个泛函是确定在所有可能的满足条件(2)的函数(1)之上的.

在薄膜的平衡位置的问题中,泛函是薄膜变形的位能 (7), 我们需要去求它在所有满足边界条件(8)的函数 $u(x, y)$ 的集合上的极小值.

每一个泛函都由两个因素来确定, 一是元素 x 的集合 M, 在 M 上这泛函被给定; 一是对每一个元素 x 赋以对应的数 (即泛函的值) 的规律. 求泛函的最大的和最小的值的方法无疑地与集合 M 的性质有关.

变分法是泛函理论的一个特殊章节. 在变分法中讨论确定在函数集合上的泛函, 而变分法的任务则是建立这种泛函的极值理论.

在建立了变分法和物理学、力学中许多部门的联系之后, 这支数学就具有特别重大的意义. 其原因可以从下面看出来. 下面即将弄明白, 为使一个函数给出泛函的极值, 这个函数必须满足一定的微分方程. 另一方面, 我们在谈微分方程的章节中已经说过,往往力学和物理学中量的规律也可以写成微分方程的形式. 而事实上许多的这种微分方程同时也是变分法中的微分方程. 这就提供了把力学的和物理学的方程考虑成相应的泛函的极值条件的可能性,和把物理学的规律表示成求某些量的极值(特别是极小值)的形式的可能性. 这就允许我们在力学和物理学中,把这样或那样的物理规律代以与之等价的"极小原理"而引进新的观点. 同时这也开辟了利用求相应泛函的极小, 来准确地或者近似地解决物理问题的新道路.

§2. 变分法的微分方程

欧拉微分方程 读者请注意，一个可微函数 f 在某点 x 具有极值的必要条件是它的微商 f' 在这点等于零：$f'(x)=0$，或者说必要条件是函数的微分等于零：$df=f'(x)dx=0$，也是一样的。

我们最近的目的是求变分法中与这相类似的条件，和弄清楚给出泛函的极值的函数应该满足怎样的必要的要求。

我们来证明，这种函数应该满足一定的微分方程。方程的形状依赖于所研究的泛函的形状。我们从所谓变分法的最简单的积分开始叙述，所谓变分法的最简单的积分就是具有下列积分表达式的泛函：

$$I(y)=\int_{x_1}^{x_2}F(x,\ y,\ y')dx. \tag{9}$$

积分符号后面的函数 F 依赖于三个自变量 $(x,\ y,\ y')$。假定当变量 y' 取任何值而变量 x 与 y 在 Oxy 平面的某区域 B 内时，函数 F 确定且对 y' 及 $x,\ y$ 两次连续可微。

令 y 为 x 的一个函数

$$y=y(x). \tag{10}$$

设它在区间 $x_1\leqslant x\leqslant x_2$ 上连续可微，而 y' 为它的导数。

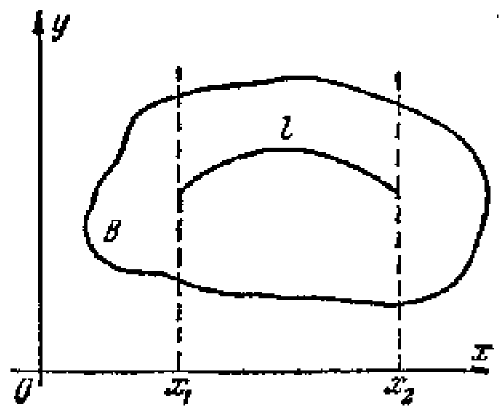


图 3

几何上，函数 $y(x)$ 是平面 Oxy 上在区间 $[x_1,\ x_2]$ 上的一条曲线 l（见图 3）。

积分 (9) 是在最快滑行曲线问题中遇到的积分 (3) 和在面积最小的旋转曲面问题中遇到的积分 (6) 的推广。积分 (9) 的值依赖于函数 $y(x)$ 或曲线 l 的选择，而关于它的极小的问题具有下列意义。

给定函数(10)（曲线 l）的集合 M。要在这个集合内找出一个

函数 y(曲线 l) 使积分 $I(y)$ 取最小值.

首先我们应该确切地定义这个函数集合 M, 在这个集合 M 上,我们来研究积分(9)的值. 在变分法中我们通常称这个集合中的函数为容许作比较的函数. 考虑固定边值的问题. 这里容许函数的集合由以下两个要求来确定:

(1) $y(x)$在区间 $[x_1, x_2]$ 上连续可微;

(2) 在区间的两端 $y(x)$ 取预先给定的值:

$$y(x_1)=y_1, \quad y(x_2)=y_2. \tag{11}$$

在其他别的方面,函数 $y(x)$ 可以完全是任意的. 如果用几何的话来说,我们研究区间 $[x_1, x_2]$ 上的所有可能的经过两点 $A(x_1, y_1)$ 及 $B(x_2, y_2)$ 且可由方程 (10) 给出的光滑曲线. 我们假定给出积分的极小的函数存在并且把它记作 $y(x)$.

下面的简单而又巧妙的想法,在变分法中经常运用着,它给出了很简单地弄清楚 $y(x)$ 所应当满足的必要条件的可能性. 主要的事情是它允许把积分(9)的极小问题转变成为函数的极小问题.

考虑依赖于数字参变量 α 的函数族

$$\bar{y}(x)=y(x)+\alpha\eta(x) \tag{12}$$

为使对于任何 α, $\bar{y}(x)$ 仍是容许函数,我们应当假定 $\eta(x)$ 连续可微并且在区间$[x_1, x_2]$的两端都等于零:

$$\eta(x_1)=\eta(x_2)=0. \tag{13}$$

就 $\bar{y}(x)$ 计算,积分(9)是参变量 α 的一个函数

$$I(\bar{y})=\int_{x_1}^{x_2}F(x,\ y+\alpha\eta,\ y'+\alpha\eta')\,dx=\phi(\alpha)^{1)}.$$

因为 $y(x)$ 给出了积分的极小值,函数 $\phi(\alpha)$ 应该在 $\alpha=0$ 时具有极小值,而其导数在这点就当然要等于零,

$$\phi'(0)=\int_{x_1}^{x_2}[F_y(x,\ y,\ y')\eta+F_{y'}(x,\ y,\ y')\eta']\,dx=0. \tag{14}$$

上面这个等式是任何连续可微并且在区间 $[x_1, x_2]$两端为零

1) 差 $\bar{y}-y=\alpha\eta$ 叫做函数 y 的变分(即改变量)并记作 δy, 而差 $I(\bar{y})-I(y)$ 叫做积分(9)的全变分. 因此就有了变分法的名称.

的函数 $\eta(x)$ 所都应当满足的．为了获得由此引出的结果，最好把条件(14)中的第二项用分部积分法变成

$$\int_{x_1}^{x_2} F_{y'}\eta' dx = -\int_{x_1}^{x_2} \eta \frac{d}{dx} F_{y'} dx,$$

而把条件(14)变成了另外的形状

$$\Phi'(0) = \int_{x_1}^{x_2} \left[F_y - \frac{d}{dx} F_{y'} \right] \eta \, dx = 0. \qquad (15)$$

可以证明下面的简单引理．

假定满足条件：

(1) 函数 $f(x)$ 在区间 $[a, b]$ 上连续；

(2) 函数 $\eta(x)$ 在区间 $[a, b]$ 上连续可微且在区间两端为零．

如果对于任何的这种函数 $\eta(x)$，积分 $\int_a^b f(x)\eta(x)dx$ 等于零，那么就有 $f(x) \equiv 0$．

事实上，假定在某一点上函数 f 不等于零，我们证明那就必然要有使 $\int_a^b f(x)\eta(x)dx \neq 0$ 的函数 $\eta(x)$ 存在，而与引理的条件矛盾．

因为 $f(c) \neq 0$ 而 f 连续，故在 C 点附近有一个区间 $[\alpha, \beta]$ 存在，在这区间内 f 处处不等于零，因而也就不变号．

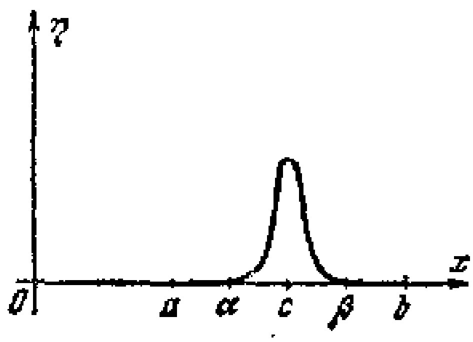


图 4

我们总可以作出函数 $\eta(x)$ 使在区间 $[a, b]$ 上连续可微，在 $[\alpha, \beta]$ 上恒正，而在 $[\alpha, \beta]$ 之外处处等于零(见图 4)．

函数 $\eta(x)$ 可以是由下列等式确定的：

$$\eta(x) = \begin{cases} 0, & \text{在}[a, \alpha]\text{上}, \\ (x-\alpha)^2(\beta-x)^2, & \text{在}[\alpha, \beta]\text{上}, \\ 0, & \text{在}[\beta, b]\text{上}. \end{cases}$$

但是对于这个函数 $\eta(x)$，

$$\int_a^b f\eta\, dx = \int_\alpha^\beta f\eta\, dx.$$

然而上面这个积分不能等于零，因为乘积 $f\eta$ 在积分区间内不为零，也不变号.

由于等式(15)是要对于任何连续可微且在区间 $[x_1, x_2]$ 两端都为零的 $\eta(x)$ 都成立的，我们根据引理，就可以肯定这只有在

$$F_y - \frac{d}{dx} F_{y'} = 0 \tag{16}$$

时才可能发生. 计算出对 x 的导数，(16)就是

$$\begin{aligned} F_y(x, y, y') - F_{xy'}(x, y, y') - F_{yy'}(x, y, y')y' \\ - F_{y'y'}(x, y, y')y'' = 0. \end{aligned} \tag{17}$$

这等式是关于函数 y 的二阶微分方程. 它叫做欧拉方程.

我们可以得出下列结论.

如果函数 $y(x)$ 给出了积分 $I(y)$ 的极小值，那么它就应当满足欧拉微分方程(17). 这在变分法中的意义和在函数的极值理论中的必要条件 $df=0$ 的意义完全一样. 它使我们马上排除了所有不满足这个条件的容许函数，因为对于这些函数积分必然不可能达到极小值. 这就非常有力地压缩了我们必须研究的容许函数的范围，于是我们可以把我们的注意力集中到方程(17)的解了.

方程(17)的解具有对于任何 $\eta(x)$ 都使导数

$$\left\{\frac{d}{d\alpha} I(y+\alpha\eta)\right\}_{\alpha=0} = 0$$

的性质，而它的意义相当于函数的逗留点. 所以往往说积分 $I(y)$ 在(17)的解上具有逗留值.

在我们的固定边值的问题中，远不必求欧拉方程的所有的解，我们只须求其中在点 x_1, x_2 取预先给定的值 y_1, y_2 的那些解.

注意欧拉方程(17)是二阶方程，它的通解要包含两个任意常数

$$y = \varphi(x, C_1, C_2).$$

这两个任意常数应当要由积分曲线通过点 A 及 B 的条件来决定. 这就给出了两个用来求常数 C_1, C_2 的方程

$$\varphi(x_1, C_1, C_2)=y_1,\ \varphi(x_2, C_1, C_2)=y_2.$$

在许多情形下这组方程只有一组解，那时就只有一条积分曲线经过 A 和 B.

寻找可能给出积分的最小值的函数的问题已化为解下列微分方程的边值问题，即在区间 $[x_1, x_2]$ 上求方程(17)的能在区间两端依次取值 y_1, y_2 的解.

后面的这个问题，往往能够利用微分方程理论中的熟知的方法来解决.

我们再一次指出，这个边值问题的解还只仅仅是具有给出最小值的可能性，在以后还必须再检查它是不是的确给出积分的最小值. 但特别在应用上常遇到的特殊情况下，欧拉方程完全解决了求积分的最小值的问题. 假定事先已经知道，给出积分的最小值的函数存在，并且还假设欧拉方程只有一个解满足边值条件(11)，因而就只有一条容许曲线能够给出最小值. 在这些条件下，可以肯定，所求得的方程(17)的解，确实给出积分的最小值.

例题 前面已经得出，最快滑行曲线的问题可以化成求积分

$$I(y)=\int_0^{x_2}\frac{\sqrt{1+y'^2}}{\sqrt{y}}\,dx$$

在满足边值条件

$$y(0)=0,\ y(x_2)=y_2$$

的函数集合上的最小值.

在这个问题里

$$F=\sqrt{\frac{1+y'^2}{\sqrt{y}}}.$$

欧拉方程为 $$-\frac{1}{2}y^{-\frac{3}{2}}\sqrt{1+y'^2}-\frac{d}{dx}\left(y^{-\frac{1}{2}}\frac{y'}{\sqrt{1+y'^2}}\right)=0.$$

经过化简之后，它可以化成

$$\frac{2y''}{1+y'^2}=-\frac{1}{y}.$$

把等式两边都乘以 y'，并积分之，得

$$\ln(1+y'^2)=-\ln y+\ln k$$

或

$$y'^2=\frac{k}{y}-1,$$

$$\sqrt{\frac{y}{k-y}}\,dy=\pm dx.$$

现在令 $y=\frac{k}{2}(1-\cos u)$, $dy=\frac{k}{2}\sin u\,du$.

经过代换与化简以后，得到

$$\frac{k}{2}(1-\cos u)\,du=\pm dx.$$

于是，积分以后即得 $x=\pm\frac{k}{2}(u-\sin u)+C$. 因为曲线应该经过坐标原点，故有 $C=0$.

由此可见，最速降线是旋轮线

$$x=\frac{k}{2}(u-\sin u),\quad y=\frac{k}{2}(1-\cos u).$$

常数 k 应当从这条曲线经过点 $M_2(x_2,\ y_2)$ 的条件求出.

依赖于几个函数的泛函 我们研究过的、变分法的最简单的泛函 (9) 只依赖于一个函数. 在应用上当研究的对象(或其行为)只由一个函数关系确定时，就会遇到这种泛函. 例如，平面上由点的横坐标与纵坐标的依赖关系来确定的曲线，由质点的坐标与时间的依赖关系来确定的该质点沿轴的运动，等等.

同样也常常可以遇到不能像这样简单地确定的对象. 为了给出空间的曲线，必须有两个函数关系来把它的两个坐标用第三个坐标表示，空间的点的运动由它的三个坐标对时间的依赖关系来确定，等等. 研究这种较复杂的对象就引出了具有几个变函数的变分问题*.

我们只限于讨论泛函依赖于两个函数 $y(x)$ 与 $z(x)$ 的情形，因为函数的个数更多时在原则上并没有什么分别.

我们提出下列问题. 容许进行比较的一对对的函数 $y(x)$ 与 $z(x)$ 由下列条件确定:

---

* 具有几个变函数的变分问题，就是具有几个函数作为变量的泛函的变分问题.——译者注

(1) 函数

$$y=y(x),\ z=z(x) \tag{18}$$

都在区间$[x_1,\ x_2]$上连续可微.

(2) 在区间两端，这些函数取已知值

$$\begin{aligned} y(x_1)=y_1,\ y(x_2)=y_2,\\ z(x_1)=z_1,\ \ z(x_2)=z_2. \end{aligned} \tag{19}$$

在所有可能的这种一对对的函数 $y(x)$ 与 $z(x)$ 中，我们要求出一对，使积分

$$I(y,\ z)=\int_{x_1}^{x_2}F(x,\ y,\ z,\ y',\ z')dx \tag{20}$$

取最小值.

在三维空间 $x,\ y,\ z$ 中，每一对容许函数对应于一条由方程(18)确定而又经过点 $M_1(x_1,\ y_1,\ z_1)$ 及 $M_2(x_2,\ y_2,\ z_2)$ 的曲线 l.

我们应当求积分(20)在所有这种曲线组成的集合上的最小值.

假定给出积分的极小值的一对函数存在，并记作 $y(x)$ 与 $z(x)$. 我们在考察这对函数的同时，我们还考察其他的函数对

$$\bar{y}=y+\alpha\eta(x),\ \bar{z}=z+\alpha\zeta(x),$$

式中 $\eta(x)$ 与 $\zeta(x)$ 是任意一对连续可微而在区间的两端 $x_1,\ x_2$ 都等于零的函数. $\bar{y},\ \bar{z}$ 于是也是容许函数，当 $\alpha=0$ 时，它们就是函数 y 和 z. 把它们代入(20)

$$\begin{aligned} I(\bar{y},\ \bar{z})&=\int_{x_1}^{x_2}F(x,\ y+\alpha\eta,\ z+\alpha\zeta,\ y'+\alpha\eta',\ z'+\alpha\zeta')dx\\ &=\Phi(\alpha), \end{aligned}$$

得到的积分是 α 的函数. 因为当 $\alpha=0$ 时，$\bar{y},\ \bar{z}$ 就是 $y,\ z$，函数 $\Phi(\alpha)$ 当 $\alpha=0$ 时应取极小值. 在取极小值的点上，Φ 的导数应等于零，

$$\Phi'(0)=0.$$

计算出导数即有

$$\int_{x_1}^{x_2}\{F_y\cdot\eta+F_z\cdot\zeta+F_{y'}\cdot\eta'+F_{z'}\cdot\zeta'\}dx=0,$$

或者，把含 η' 及 ζ' 的两项分部积分以后，有

$$\int_{x_1}^{x_2}\left\{\left[F_y-\frac{d}{dx}F_{y'}\right]\eta(x)+\left[F_z-\frac{d}{dx}F_{z'}\right]\zeta(x)\right\}dx=0.$$

这个等式应该对于任何一对连续可微且在区间两端都等于零的函数 $\eta(x)$ 与 $\zeta(x)$ 都满足．所以，根据前面证明过的基本引理，容易断言下列两个条件应当满足：

$$\begin{aligned}&F_y-\frac{d}{dx}F_{y'}=0,\\&F_z-\frac{d}{dx}F_{z'}=0.\end{aligned}\tag{21}$$

所以，如果函数 y, z 使积分 (20) 取极小值，它们就应当满足欧拉微分方程组(21)．

这种结论再一次允许我们把关于积分(20)的极小值的变分问题代以微分方程论的边值问题：试在区间 $[x_1, x_2]$ 上，求微分方程组(21)的满足边值条件(19)的解 y, z.

和前面的情形一样，这开辟了解决所提出的极小值问题的一个可能的途径．

作为应用欧拉方程组 (21) 的例子，我们考察牛顿力学中的奥斯特洛格拉特斯基-汉弥登变分原理，我们只考察这个原理的最简单的形式．

取质量为 m 的物体，并且认为这物体的大小与形状可以忽略而可以把它看成一个质点．

假定质点从它在时间 t_1 所在的位置 $M_1(x_1, y_1, z_1)$，在时间 t_2 移到了位置 $M_2(x_2, y_2, z_2)$．假定运动服从牛顿力学的规律，并且运动是由于力 $\boldsymbol{F}(x, y, z, t)$ 的作用产生的，这力依赖于点的位置和时间 t 并且具有位函数 $U(x, y, z, t)$．所谓力 $\boldsymbol{F}$ 具有位函数 U 的含义是力 $\boldsymbol{F}$ 在三个坐标轴方向的分量 F_x, F_y, F_z 是某一个函数 U 对于相应的坐标的偏导数：

$$F_x=\frac{\partial U}{\partial x},\quad F_y=\frac{\partial U}{\partial y},\quad F_z=\frac{\partial U}{\partial z}.$$

假定运动是自由的，不受任何约束[1]。

牛顿的运动方程是

$$m\frac{d^2x}{dt^2}=\frac{\partial U}{\partial x},\quad m\frac{d^2y}{dt^2}=\frac{\partial U}{\partial y},\quad m\frac{d^2z}{dt^2}=\frac{\partial U}{\partial z}.$$

根据牛顿力学的规律，点按确定的途径完成位移。与点的"牛顿运动"的同时，我们考察它的其他运动，这些运动简称"容许的运动"。我们用两个要求来定义"容许的运动"，即在时间 t_1 质点据有位置 M_1 而在时间 t_2 则据有位置 M_2。

怎样来区别质点的"牛顿运动"和它的任何其他的"容许的运动"呢？奥斯特洛格拉特斯基-汉弥登原理就给出了区别这个的可能性。

引进质点的动能

$$T=\frac{1}{2}m(x'^2+y'^2+z'^2)$$

并作出所谓作用积分

$$I=\int_{t_1}^{t_2}(T+U)dt.$$

原理的内容是：质点的"牛顿运动"与它的其他"容许的运动"的区别就在于"牛顿运动"使作用积分达到逗留值。

作用积分依赖于三个函数：$x(t)$，$y(t)$，$z(t)$。

由于在所有进行比较的运动中，质点的开始位置和终了位置都是一样的，这些函数的边值是固定的。这里我们就有了具有三个在区间 $[t_1, t_2]$ 的端点取固定的值的变函数的变分问题。

前面曾经约定，如果某条曲线是欧拉方程的积分曲线，我们就说在这曲线上积分(17)具有逗留值。在我们的问题中，被积函数

$$F=T+U=\frac{1}{2}m(x'^2+y'^2+z'^2)+U(x, y, z, t),$$

依赖于三个函数，而对于积分的逗留值，应该满足三个微分方程组

1) 对于奥斯特洛格拉特斯基-汉弥登原理这是非本质的：在力学体系上可以添加任何的，甚至是非定常的但只要是整式的约束，也就是说，可以写成方程式的形式而不包含坐标对于时间的导数的约束。

成的方程组

$$F_x - \frac{d}{dt} F_{x'} = 0,$$

$$F_y - \frac{d}{dt} F_{y'} = 0,$$

$$F_z - \frac{d}{dt} F_{z'} = 0.$$

由于 $F_x = \frac{\partial U}{\partial x}$, $F_{x'} = mx'$, … 等等，欧拉方程组就和牛顿力学中的运动方程组相同．这就建立了上述原理的正确性．

多重积分的极小值问题 我们要求读者注意的最后一个变分法的问题是多重积分的极小值问题．由于和解决这种问题有关的事实对于任何重数的多重积分都是类似的，我们只就多重积分中最简单的情形即二重积分来讨论．

设 B 为平面 Oxy 上由围线 l 围成的区域． 容许作比较的函数集合由下列条件确定：

(1) $u(x, y)$ 在区域 B 内连续可微；

(2) u 在 l 上取给定的值

$$u|_l = f(M). \tag{22}$$

在所有函数 u 中，求使积分

$$I(u) = \iint_B F(x, y, u, u_x, u_y)\,dx\,dy \tag{23}$$

取极小值的函数．

函数 u 的边界值的给定，意味着在空间 (x, y, u) 中给定了在 l 之上的一条空间围线 Γ(参看第 161 页图 2)．

考虑所有经过 Γ 而在区域 B 上的一切可能的曲面 S． 在它们之中，我们希望求出使积分(23)为极小的曲面．

像以前一样，我们假定给出积分的极小的函数存在，并且把它记做 u. 同时考虑其他的函数

$$\bar{u} = u + \alpha\eta(x, y)$$

式中 $\eta(x, y)$ 是任何连续可微函数，并且在 l 上等于零．于是函数

$$I(u)=\iint_B F(x,\ y,\ u+\alpha\eta,\ u_x+\alpha\eta_x,\ u_y+\alpha\eta_y)\,dx\,dy=\Phi(\alpha)$$

应当在 $\alpha=0$ 时取极小值. 在这种情形下，它的一阶导数在 $\alpha=0$ 时应为零，

$$\Phi'(0)=0,$$

或

$$\iint_B [F_u\eta+F u_x\eta_x+F u_y\eta_y]\,dx\,dy=0. \tag{24}$$

我们用奥斯特洛格拉特斯基公式来改变上式中后面的两项.

$$\begin{aligned}\iint_B &[F u_x\eta_x+F u_y\eta_y]\,dx\,dy\\ &=\iint_B\left[\frac{\partial}{\partial x}(F u_x\eta)+\left[\frac{\partial}{\partial y}(F u_y\eta)\right]dx\,dy\right.\\ &\quad-\iint_B\left[\frac{\partial}{\partial x}(F u_x)+\frac{\partial}{\partial y}(F u_y)\right]\eta\,dx\,dy\\ &=\int_l[F u_x\cos(n,\ x)+F u_y\cos(n,\ y)]\,\eta\,ds\\ &\quad-\iint_B\left[\frac{\partial}{\partial x}(F u_x)+\frac{\partial}{\partial y}(F u_y)\right]\eta\,dx\,dy.\end{aligned}$$

沿 l 的围线积分应该等于零，因为在围线 l 上函数 η 等于零. 所以条件(24)可以表成形式

$$\iint_B\left[F_u-\frac{\partial}{\partial x}(F u_x)-\frac{\partial}{\partial_y}(F u_y)\right]\eta\,dx\,dy=0.$$

这个等式应该为任何连续可微而在边界 l 上为零的函数 η 所满足.

所以，和前面的类似，可以断定，在区域 B 的所有的点上，应当满足方程

$$F_u-\frac{\partial}{\partial x}F u_x-\frac{\partial}{\partial y}F u_y=0. \tag{25}$$

于是，如果函数 u 给出积分 (23) 的极小值，它就应当满足偏微分方程(25).

正像所有前面的问题一样，这里也建立了求一个积分的极小值的变分法的问题与微分方程(在这里是偏微分方程)的边值问题的联系.

例题 边界上有变形的薄膜上的点的偏差 $u(x, y)$ 应当从在给定边值 $u|_l=\varphi$ 时位能

$$\frac{\mu}{2}\iint_B [u_x^2+u_y^2]\,dx\,dy$$

是极小的条件求出.

为简单计,不考虑常数因子 μ, 可以取

$$F=\frac{1}{2}(u_x^2+u_y^2),$$

而方程(25)具有形状

$$-\frac{\partial}{\partial x}(u_x)-\frac{\partial}{\partial y}(u_y)=0,$$

或

$$\Delta u=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0.$$

因此, 确定薄膜的点的偏差的问题化成了求在区域的围线上取给定值 φ 的调和函数了(参看第 6 章, § 3).

§ 3. 变分法问题的近似解法

我们指出变分法问题一些近似解法的思想来结束这一章.

为确定计,我们只对固定边值的容许函数讨论最简单的泛函

$$I(y)=\int_{x_1}^{x_2}F(x, y, y')\,dx.$$

假定 $y(x)$ 是求 I 的极小的问题的解, 而 $m=I(y)$ 是积分的相应的极小值. 显然,如果我们能求出容许函数 $\bar{y}$, 使积分的对应值 $I(\bar{y})$ 和 M 很接近, 那么我们可以预期 $\bar{y}$ 也和准确的解 y 相差很小. 此外,如果我们做出了一序列容许函数 $\bar{y}_1, \bar{y}_2, \cdots$, 使 $I(\bar{y}_n)\to m$, 那么可以希望这个序列会按这种或那种意义收敛到解 y, 而对于充分大的下标计算出 $\bar{y}_n$, 我们就可以求出具有任意高的准确度的解.

根据我们所选择建立"极小化序列" $\bar{y}_n(n=1, 2, \cdots)$ 的这样或那样的法则,我们就得到了变分问题的这种或那种近似解法.

最早的近似方法是折线法或欧拉法．把区间 $[x_1, x_2]$ 分成许多部分．例如，把这些部分取成一样长，使分点为

$$x_1,\ x_1+h,\ x_1+2h,\ \cdots,\ x_1+nh=x_2,\ h=\frac{x_2-x_1}{n}.$$

做出顶点在分点之上的折线 p_{n-1}．它的顶点的纵坐标记做

$$b_0,\ b_1,\ b_2,\ \cdots,\ b_{n-1},\ b_n.$$

并且要求这些折线具有相同的起点与终点．这样所有的容许折线都合于 $b_0=y_1$, $b_n=y_2$．所以折线由纵坐标

$$b_1,\ b_2,\ \cdots,\ b_{n-1}$$

确定．

现在的问题就是应该如何选择折线 p_{n-1}（也就是选择它的顶点 b_i 的纵坐标）使它能尽可能的最接近问题的准确解．

为了达到这个目的，自然这样来进行．对折线计算积分 I 的值，这个值要依赖于 b_i:

$$I(p_{n-1})=\Phi(b_1,\ b_2,\ \cdots,\ b_{n-1}),$$

也就是说，积分是这些纵坐标的函数．现在选择 b_i 使 $I(p_{n-1})$ 取极小值．为了确定所有的 b_i，我们有方程组

$$\frac{\partial}{\partial b_i}I(p_{n-1})=0,\ (i=1,\ 2,\ \cdots,\ n-1).$$

因为对于任何容许曲线，包括问题的准确解，都可以用这样的折线去任意地逼近，使得不但它们在平面上的位置很接近，而切线的方向也可以很接近，所以非常明显，从计算结果就得出一序列的折线 p_{n-1}，这个序列差不多肯定将是极小化序列．取 n 充分大，我们可求出在整个区间 $[x_1,\ x_2]$ 上和解近似到任何程度的折线．当然，收敛性的事情是应当对每一个情形进行具体研究的．

下面的非常便于计算的方法也在物理学和技术科学中极广泛地应用．

取一个满足边值条件 $\varphi_0(x_1)=y_1$, $\varphi_0(x_2)=y_2$ 的任意函数 $\varphi_0(x)$ 及一串在区间 $[x_1,\ x_2]$ 两端为零的函数 $\varphi_1(x)$, $\varphi_2(x)$, $\cdots$.

然后作线性组合

$$S_n(x)=\varphi_0(x)+a_1\varphi_1(x)+\cdots+a_n\varphi_n(x).$$

对于系数 a_1, a_2, …, a_n 的任何一组数值，函数 $S_n(x)$ 都是容许函数.

用 $S_n(x)$ 代替积分 I 中的 y, 并完成必要的计算,我们就得到系数 a_i 的一个函数.

现在选择 a_i 使这个函数取它的最小值,这些系数应从方程

$$\frac{\partial}{\partial a_i}I(S_n)=0 \quad (i=1, 2, \cdots, n)$$

求出. 解这组方程,一般来说,我们得到给出 $I(S_n)$ 的极小值的确定的一组系数值 $\bar{a}_1$, …, $\bar{a}_n$, 从而作出了解的近似式

$$\bar{S}_n(x)=\varphi_0(x)+\bar{a}_1\varphi_1(x)+\cdots+\bar{a}_n\varphi_n(x).$$

这种近似序列 $\bar{S}_n(n=1, 2, \cdots)$, 不是对于任何选择的 φ_i 都是极小化序列. 为了使得 $\bar{S}_n$ 是极小化序列, 函数序列 φ_i 必须满足某些"完备性"条件,关于这些的说明我们就不再涉及了..

文　　献

Л. А. Люстерник, Кратчайшие линии (вариационные задачи), Популярные лекции по математике, вып. 19. Гостехиздат, 1956.

大 学 教 材

Н. И. Ахиезер, Лекции по вариационному исчислению, Гостехиздат, 1955.

М. А. Лаврентьев и Л. А. Люстерник, Курс вариационного исчисления, Гостехиздат, 1950.

В. И. Смирнов, Курс высшей математики, т. IV, изд. 3-е, Гостехиздат, 1953.

胡祖炽 译

冯　康 校

第九章 复变函数

§1. 复数和复变函数

复数和它在代数中的意义 在数学中，由于解代数方程的关系，我们就引进了复数。代数方程

$$x^2+1=0 \tag{1}$$

在实数域中是不可解的，因此，就促使我们引入一个由等式

$$i^2=-1 \tag{2}$$

所定义的约定的数 i，叫做虚单位。设 a 和 b 是实数，则形如 $a+bi$ 的数叫做**复数**。关于这种数的运算，我们把它们当作二项式来加、来乘，正如同运算实数一样。假若这时我们利用等式 (2)，则复数经过算术运算，结果仍旧得到复数[1]。复数的除法运算，这是定义作乘法的逆运算的，也是一个经常成立的一意运算，只要除数不等于零就行。这样一来，复数的引进首先就说明了一个值得注意的(但暂时还是形式上的)现象：除了实数之外，还存在另外的数，关于复数，所有的算术运算皆成立。

下一步就是复数的几何解释。每一复数 $a+bi$ 可以用 Oxy 平面上坐标为(a, b)的点来表示，或者用一个从原点引到(a, b)点的向量来表示，这对复数又引进了一个新的看法。复数，这是一个实数对 (a, b)，我们对它，已经明确地规定了加法运算和乘法运算，它们和实数运算一样服从同样的规律。这时发生了一个值得注意的现象：两个复数之和

$$(a+bi)+(c+di)=(a+c)+(b+d)i$$

在几何上由表示两个被加项的向量作成的平行四边形的对角线来表示(图 1)。因此，复数是按照力学和物理学中所出现的向量：力、

1) 读者在中学里就应当已经知道了复数，参考第四章(第一卷)§3.

速度、加速度等相加的规律来相加的. 这就使得我们有理由认为复数不仅具有单纯形式推广的意义，而且还可以用来表示实有的物理量.

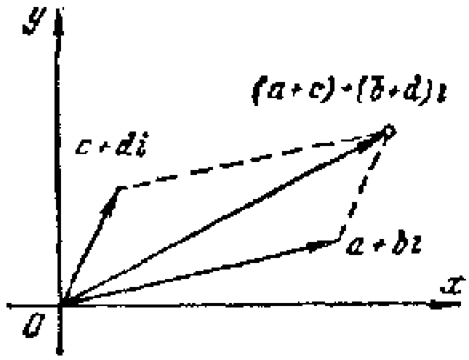


图 1

我们进一步还可以看出，这种看法在实际的数学物理问题中将取得多么大的成就.

然而把复数引进来这一件事首先还是在发现代数和分析的规律方面开始带来了成就. 实数域，它对于算术运算说来是封闭的，但对于代数运算则是不够完全的，就像(1)这样一个简单的方程在实数域中尚且没有根. 高等代数的基本定理，即任何复系数代数方程

$$z^n+a_1z^{n-1}+\cdots+a_{n-1}z+a_n=0$$

必有 n 个根，这是一个特别重要的事情[1).

这定理指出，复数做成了一个数系，它就某种意义来说，关于代数运算是完全的，只把 (1) 这样一个方程的根加到实数域中去，由于导出了数 $a+bi$，在这种数做成的域中，就使得任何代数方程都可解，这不是一件简单的事情. 高等代数的基本定理已经指出了，多项式的理论，即使这些多项式都具有实系数，当且仅当我们把多项式的值在整个复数域中来考虑的时候，才可能得到最终的形式. 代数多项式论的进一步的发展越来越肯定了这种观点. 只有当我们把多项式作为一个复变函数来考虑的时候，它的性质才弄得清楚.

幂级数和复变函数 分析的发展显示出了一系列的事实，这些事实指出了，引进复数这一件事不只是在多项式论中具有重要意义，而且对于另一个非常重要的函数类，即可以展成幂级数

$$f(x)=a_0+a_1(x-a)+a_2(x-a)^2+\cdots \tag{3}$$

的函数类，也具有重大的意义.

正如我们在第二章(第一卷)中曾经说过的，无穷小量分析的发

1) 参看第四章(第一卷)§3.

展要求我们在函数这一概念上，以及在数学中定义函数的各种可能的方法上，需要建立起更为精确的观点．在这里，我们不来谈论这些富有趣味的问题，我们只提醒一下，在分析学发展的最初阶段中，它显示出了，最常遇到的函数在它的定义域之内每一点的附近皆可展成幂级数．例如所谓的初等函数就都具有这种性质．

分析中大量的具体问题皆导致可展成幂级数的函数．另一方面，人们有着一种愿望，要把"数学"函数的定义和"数学"公式联系在一起，而用来作为一个"数学"公式，幂级数则有一种适于各种用途的形式．这种现象甚至使得我们产生了一种奢望，要想把分析限于研究可以展成幂级的函数，这种函数因而得到解析函数之名．科学的发展证明了这种限制是不适当的．数学物理中的一些问题已经使得我们必须跑到解析函数类的范围外面去．解析函数类就连（比如说）那些可以用具有角点的曲线来表示的函数也不包括．但解析函数类，由于它的值得注意的性质以及它的多方面的用途，它已经成为数学中所研究的最重要的一类函数．

因为在计算幂级数的每一项时，只需要用到算术运算，所以那些可以用幂级数来表示的函数的函数值对于变数取复数值时也可以算出，只要对于这些值级数仍然保证收敛就行，像这样定义实变函数关于复变数的值，我们称之为它在复域内的"扩展"．因此，解析函数，正如多项式一样，不仅对于实变数值可以进行讨论，而且对于复变数值也可以进行讨论．不仅如此，我们还可以研究复系数的幂级数，解析函数的性质，正如多项式的性质一样，当且仅当把它放在复域中来讨论的时候，才能完全予以揭露．我们现举一例阐明上面所说．

我们现在来看两个实变函数

$$e^x \quad 和 \quad \frac{1}{1+x^2}.$$

这两个函数在整个实轴 Ox 上都是有限的、连续的，而且可以微分任意多次，它们可以展成泰勒级数，例如在原点 $x=0$ 的附近，有

$$e^x=1+\frac{x}{1!}+\frac{x^2}{2!}+\cdots, \tag{4}$$

$$\frac{1}{1+x^2}=1-x^2+x^4-x^6+\cdots. \tag{5}$$

所得的级数前一个对于一切 x 值皆为收敛，而第二个级数则只当 $-1<x<+1$ 时收敛．若仅就变数取实值时来研究函数(5)，则不可能发现出到底是它的什么样的性质使得它的泰勒级数在 $|x|\geqslant 1$ 时发散．但若转到复域中来看，则即可说明这一现象，我们现在对于变数取复值时来研究级数(5)．

$$1-z^2+z^4-z^6+\cdots \tag{6}$$

这级数的 n 项之和

$$s_n=1-z^2+z^4-z^6+\cdots+(-1)^{n-1}z^{2n-2},$$

可以如同对于实值 z 一样算出

$$s_n+z^2s_n=1+(-1)^n z^{2n},$$

由此即得

$$s_n=\frac{1+(-1)^n z^{2n}}{1+z^2}.$$

这式子指出，当 $|z|<1$ 时，

$$\lim_{n\to\infty} s_n=\frac{1}{1+z^2},$$

因为 $|z|^{2n}\to 0$．由是，对于满足不等式 $|z|<1$ 的复数 z，级数 (6) 收敛，并以 $\frac{1}{1+z^2}$ 为其和，当 $|z|\geqslant 1$ 时，级数 (6) 发散，因为这时差数 $s_n-s_{n-1}=(-1)^{n-1}z^{2n-2}$ 不趋于零．

不等式 $|z|<1$ 指出，z 点与坐标轴原点的距离不超过 1．因此，使得级数(6)收敛的点在复平面上做成一个以坐标轴原点为心的圆，在这圆的圆周上有两个点 i 和 $-i$，函数 $\frac{1}{1+z^2}$ 在这两点变为无穷；这两点的出现也就是级数 (6) 的收敛域受到限制的原故．

幂级数的收敛域 幂级数

$$a_0+a_1(z-a)+a_2(z-a)^2+\cdots+a_n(z-a)^n+\cdots \tag{7}$$

在复平面上的收敛域常是一个以 a 点为心的圆．

我们现在来证明这一命题，它具有阿贝尔定理之名．

我们先来指出，复数项级数

$$w_1+w_2+\cdots+w_n+\cdots \tag{8}$$

可以看作两个级数，各由数 $w_n=u_n+iv_n$ 的实部分和虚部分的系数所组成：

$$u_1+u_2+\cdots+u_n+\cdots, \tag{9}$$

$$v_1+v_2+\cdots+v_n+\cdots. \tag{10}$$

级数(8)的部分和 s_n 可以用级数(9)和(10)的部分和 σ_n 和 τ_n 来表示：

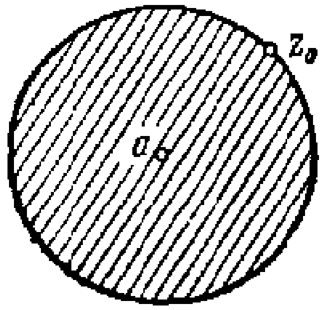


图 2

$$s_n=\sigma_n+i\tau_n,$$

因此，级数(8)的收敛与否就和级数(9)和(10)的收敛与否等价，而且级数(8)的和数 s 可以用级数(9)和(10)的和数 σ 和 τ 表出：

$$s=\sigma+i\tau.$$

在作了上述这一说明之后，下面的引理就变得很清楚。

若级数(8)的项就绝对值的大小而言小于几何级数

$$A+Aq+\cdots+Aq^n+\cdots$$

的项，此时，A 和 q 为正数，且 $q<1$，则级数(8)收敛。

事实上，若 $|w_n|<Aq^n$，则

$$|u_n|\leqslant|w_n|<Aq^n,\quad |v_n|\leqslant|w_n|<Aq^n,$$

故[参看第二章(第一卷)§ 14] 级数(9)和(10)收敛，因而级数(8)也收敛。

我们现在来证明，若幂级数(7)在某一点 z_0 收敛，则它在以 a 点为心，边界经过 z_0 的圆内亦为收敛(图 2)。假若这点已经证明，我们由此就容易推出，级数(7)

$$a_0+a_1(z-a)+\cdots+a_n(z-a)^n+\cdots$$

的收敛域或为全平面，或为 $z=a$ 一点，或为某一半径为有限的圆。

于是，设若级数(7)在 z_0 点收敛，则级数(7)的一般项对于 $z=z_0$ 当 $n\to\infty$ 时趋于零，而这也就是说，级数(7)的一切项皆在某一圆之内；设 A 为这样的一个圆的半径，则对于任何 n，

$$|a_n(z_0-a)^n|<A. \tag{11}$$

我们现在任取一点 z, 它和 a 点的距离小于 z_0 点到 a 点的距离, 而来证明级数在 z 点收敛.

显而易见,

$$|z-a|<|z_0-a|,$$

因此

$$q=\frac{|z-a|}{|z_0-a|}<1. \tag{12}$$

我们来估计级数(7)在 z 点的一般项,

$$|a_n(z-a)^n|=\left|a_n(z_0-a)^n\left(\frac{z-a}{z_0-a}\right)^n\right|$$

$$=|a_n(z_0-a)^n|\left(\frac{|z-a|}{|z_0-a|}\right)^n;$$

由不等式(11)和(12), 即得

$$|a_n(z-a)^n|<Aq^n,$$

这也就是说, 级数 (7) 在 z 点的一般项小于一收敛几何级数的一般项. 根据上面所证明的引理, 级数(7)在 z 点收敛.

一个圆, 假若幂级数于其中为收敛, 于其外为发散, 则称之为**收敛圆**. 这圆的半径叫做该级数的**收敛半径**. 收敛圆周必然通过复平面上使得函数的正则情况受到破坏的点中与 a 点最为接近的一点.

幂级数(4)在全复平面上皆为收敛; 幂级数(5), 正如已经证明了的, 它的收敛半径等于 1.

复变指数函数和三角函数 幂级数可以用来作实变函数在复平面上的"拓展". 例如对于复数值 z, 可以用下面的幂级数定义函数 e^z:

$$e^z=1+\frac{z}{1!}+\frac{z^2}{2!}+\cdots. \tag{13}$$

同理, 我们引入复变三角函数

$$\sin z=\frac{z}{1!}-\frac{z^3}{3!}+\frac{z^5}{5!}-\cdots, \tag{14}$$

$$\cos z=1-\frac{z^2}{2!}+\frac{z^4}{4!}-\cdots. \tag{15}$$

这几个级数在全平面上皆为收敛.

我们现在把三角函数和指数函数之间当转入复域时所出现的关系予以指出,这将是有益处的.

在等式(13)中以 iz 代 z, 即得

$$e^{iz}=1+i\frac{z}{1!}-\frac{z^2}{2!}-i\frac{z^3}{3!}+\frac{z^4}{4!}+\cdots$$

将无因子 i 的各项和有因子 i 的各项分别集中在一起,则得

$$e^{iz}=\cos z+i\sin z. \tag{16}$$

同理可得

$$e^{-iz}=\cos z-i\sin z. \tag{16'}$$

公式(16)和(16′)名为欧拉公式. 关于 $\cos z$ 和 $\sin z$ 解出(16)和(16′), 即得

$$\cos z=\frac{e^{iz}+e^{-iz}}{2},$$

$$\sin z=\frac{e^{iz}-e^{-iz}}{2i}. \tag{17}$$

非常重要的是,指数定律对于复变数值仍然成立:

$$e^{z_1}\cdot e^{z_2}=e^{z_1+z_2}. \tag{18}$$

因为对于复变数值我们已经用级数 (13) 定义了函数 e^z, 故(18)式应当由这定义出发来加以证明.

由定义,

$$e^{z_1}\cdot e^{z_2}=\left(1+\frac{z_1}{1!}+\frac{z_1^2}{2!}+\cdots\right)\cdot\left(1+\frac{z_2}{1!}+\frac{z_2^2}{2!}+\cdots\right).$$

我们现在把级数逐项相乘. 在乘出级数时, 我们可以将所得之项写成正方形:

$$\begin{array}{l}
1\cdot 1+1\cdot\frac{z_2}{1!}+1\cdot\frac{z_2^2}{2!}+1\cdot\frac{z_2^3}{3!}+\cdots\\
\cdots+\frac{z_1}{1!}\cdot 1+\frac{z_1}{1!}\cdot\frac{z_2}{1!}+\frac{z_1}{1!}\cdot\frac{z_2^2}{2!}+\frac{z_1}{1!}\cdot\frac{z_2^3}{3!}+\cdots\\
\cdots+\frac{z_1^2}{2!}\cdot 1+\frac{z_1^2}{2!}\cdot\frac{z_2}{1!}+\frac{z_1^2}{2!}\cdot\frac{z_2^2}{2!}+\frac{z_1^2}{2!}\cdot\frac{z_2^3}{3!}+\cdots
\end{array}$$

$$\cdots+\frac{z_1^3}{3!}\cdot 1+\frac{z_1^3}{3!}\cdot\frac{z_2}{1!}+\frac{z_1^3}{3!}\cdot\frac{z_2^2}{2!}+\frac{z_1^3}{3!}\cdot\frac{z_2^3}{3!}+\cdots.$$

我们现在把同类项集合在一起. 容易证明，在我们的表上这种项皆在同一对角线上. 于是即得

$$e^{z_1}\cdot e^{z_2}=1+\left(\frac{z_2}{1!}+\frac{z_1}{1!}\right)+\left(\frac{z_2^2}{2!}+\frac{z_2}{1!}\cdot\frac{z_1}{1!}+\frac{z_1^2}{2!}\right)+\cdots. \quad (19)$$

这级数的一般项为

$$\frac{z_2^n}{n!}+\frac{z_2^{n-1}}{(n-1)!}\frac{z_1}{1!}+\frac{z_2^{n-2}}{(n-2)!}\frac{z_1^2}{2!}+\cdots+\frac{z_1^n}{n!}$$
$$=\frac{1}{n!}\left(z_2^n+\frac{n!}{1!(n-1)!}z_2^{n-1}z_1+\frac{n!}{2!(n-2)!}z_2^{n-2}z_1^2+\cdots+z_1^n\right).$$

运用牛顿二项定理，则可将此一般项化为

$$\frac{(z_1+z_2)^n}{n!}.$$

于是，级数 (19) 的一般项与 $e^{z_1+z_2}$ 的一般项一致，而这也就证明了指数定律(18).

指数定律和欧拉公式使得我们可以把函数 e^z 用实变函数表成有限形式(不是级数). 实际上，设

$$z=x+iy,$$

则得
$$e^z=e^{x+iy}=e^x\cdot e^{iy},$$
又因
$$e^{iy}=\cos y+i\sin y,$$
故得
$$e^z=e^x(\cos y+i\sin y). \quad (20)$$

上面所得到的公式对于研究函数 e^z 的性质非常方便. 我们现在指出它的两个性质：(1) 函数 e^z 无一处为零；实际上，$e^x\neq 0$，而且公式 (20) 中的 $\cos y$ 和 $\sin y$ 不可能同时为零；(2) 函数 e^z 具有周期 $2\pi i$，即

$$e^{z+2\pi i}=e^z.$$

上式可从指数定律和等式

$$e^{2\pi i}=\cos 2\pi+i\sin 2\pi=1$$

推出. 公式 (17) 使得我们可以在复数域内来研究函数 $\cos z$ 和

$\sin z$. 我们留给读者去证明，在复数域内 $\cos z$ 和 $\sin z$ 皆具有周期 2π，且和数的正弦定理和余弦定理皆成立。

复变函数的一般概念，函数的可微分性 幂级数可以用来定义解析复变函数。但值得注意的是对任意一个复变函数去研究分析的基本运算，首先是微分运算。在这里出现了一个与复变函数的可微分性有关的深刻的事实。正如我们下面将要看到的，一方面，在 z_0 点附近所有的点皆具有一阶导数的函数在 z_0 点必然具有各阶导数，不仅如此，在这点还可展成幂级数，即为解析的。于是，在研究可微分的复变函数时，我们重新又回到解析函数上去。另一方面，导数的研究还使我们发现了复变函数的解析性质和函数论以及数学物理方程问题之间的关系。

根据上面所说，以后我们将把在 z_0 点的某一邻域之内的一切点皆具有导数的那种函数叫做在 z_0 点解析的函数。

从函数的一般定义出发，假若根据已给的复数值 z，我们可以找出一种规律，使得可以得出复数值 w，我们就说 w 是复数 z 的函数。

每一复数 $z=x+iy$ 可以用平面 Oxy 上的点 (x, y) 来表示，而数 $w=u+iv$ 我们则将用 Ouv 平面（函数平面）上的点表示。于是，从几何的观点看，复变函数

$$w=f(z)$$

在变数平面 Oxy 上的点与函数平面 Ouv 上的点之间定义了一个对应关系。换言之，复变函数给变数平面到函数平面上建立了一个映象。给了一个复变函数，这就是说，在数对 (x, y) 和 (u, v) 之间给了一个对应，因之，给定了一个复变函数和给定了两个函数

$$u=\varphi(x, y), \quad v=\psi(x, y)$$

是等价的，于此显然有

$$w=u+iv=\varphi(x, y)+i\psi(x, y).$$

譬如说，假设 $\quad w=z^2=(x+iy)^2=x^2-y^2+2ixy,$

则 $\quad u=\varphi(x, y)=x^2-y^2, \quad v=\psi(x, y)=2xy.$

复变函数的导数是在形式上如同实变函数的导数一样给以定

义的．导数是函数的增量与变数的增量之比的极限，

$$f'(z)=\lim_{\Delta z\to 0}\frac{f(z+\Delta z)-f(z)}{\Delta z}, \tag{21}$$

假若这极限存在的话．

若我们假定构成函数 $w=f(z)$ 的两个实函数 u 和 v 关于 x 和 y 具有偏导数，则此仍不足以保证函数 $f(z)$ 的导数存在．增量之比的极限一般是与点 $z'=z+\Delta z$ 接近点 z 的方向有关(图 3)．要想导数 $f'(z)$ 存在，必须这极限与 z' 接近 z 的方法无关．例如我们现在来考虑 z' 沿 Ox 轴的方向和 Oy 轴的方向接近于 z 这两种情形．

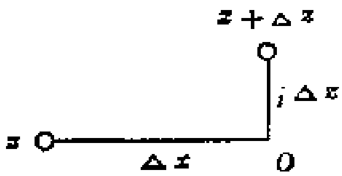


图 3

在前一种情形，

$$\Delta z=\Delta x,$$

$$f(z+\Delta z)-f(z)$$
$$=u(x+\Delta x,\ y)-u(x,\ y)+i[v(x+\Delta x,\ y)-v(x,\ y)],$$

而增量的比

$$\frac{f(z+\Delta z)-f(z)}{\Delta z}$$
$$=\frac{u(x+\Delta x,\ y)-u(x,\ y)}{\Delta x}+i\frac{v(x+\Delta x,\ y)-v(x,\ y)}{\Delta x}$$

当 $\Delta x\to 0$ 时趋于

$$\frac{\partial u}{\partial x}+i\frac{\partial v}{\partial x}. \tag{22}$$

在第二种情形，

$$\Delta z=i\Delta y,$$

而增量的比

$$\frac{f(z+\Delta z)-f(z)}{\Delta z}$$
$$=-i\frac{u(x,\ y+\Delta y)-u(x,\ y)}{\Delta y}+\frac{v(x,\ y+\Delta y)-v(x,\ y)}{\Delta y}$$

之极限为

$$\frac{\partial v}{\partial y}-i\frac{\partial u}{\partial y}. \tag{23}$$

假若函数 $w=f(z)$ 具有导数，则所得的这两个式子应该相等，因而有

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}, \tag{24}$$

$$\frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}.$$

这两个方程的成立是函数 $w=u+iv$ 的导数存在的必要条件. 原来，方程组 (24) 不仅是必要条件，而且也是充分条件(假若函数 u, v 具有全微分的话). 我们现在不来谈论条件 (24) 之为充分条件的证明. 条件(24)叫做**柯西-黎曼方程**.

容易证明，关于实变函数的一系列的微分规则可以不加改变地搬到复变函数上来. 像函数 z^n，和数、乘积、商数等的微分就是如此. 这些公式的推演和在实变函数的情形一样，不过现在指的是复量，不是实量. 这证明了，z 的一切多项式

$$w=a_0+a_1z+\cdots+a_nz^n$$

是处处可微分的函数. 任何有理函数，它等于两个多项式之比，

$$w=\frac{a_0+a_1z+\cdots+a_nz^n}{b_0+b_1z+\cdots+b_mz^m},$$

在所有使得分母不为零的点皆为可微分的.

要想证明函数 $w=e^z$ 可以微分，我们可以利用柯西-黎曼条件. 在现在这种情形下，根据公式(20)，

$$u=e^x\cos y,\ v=e^x\sin y;$$

将这两个函数代入(24)，即可证明柯西-黎曼条件成立. 这函数的导数可以(譬如)根据公式(22)算出. 由此即得

$$\frac{dw}{dz}=e^z.$$

根据公式 (17) 容易证明三角函数皆可微分，而且在分析中关于这些函数的导数的值所常遇到的那些公式仍然适用.

函数 Lnz 在这里，我们将不把所有的初等复变函数逐一来进行研究. 但知道一下函数 Lnz 的性质对我们还是非常重要的.

正如在实数域中一样，若

$$z=e^{w},$$

则令 $w=Lnz$.

为了把函数 Lnz 加以分析，我们现在把数 z 写成三角形式

$$z=r(\cos\varphi+i\sin\varphi).$$

利用指数律于 e^{w}，则得

$$z=e^{w}=e^{u+iv}=e^{u}\cdot e^{iv}=e^{u}(\cos v+i\sin v).$$

将所得的关于 z 的两个式子加以比较，即得

$$e^{u}=r, \tag{α}$$

$$\cos v+i\sin v=\cos\varphi+i\sin\varphi. \tag{β}$$

注意 u 和 r 为实数，则由公式(α)，即得

$$u=Ln\,r,$$

于此，$Ln\,r$ 是实数的自然对数的普通值．等式(β)仅当

$$\cos v=\cos\varphi,\ \sin v=\sin\varphi$$

时成立，而对于这等式，v 与 φ 只能相差一 2π 的倍数，

$$v=\varphi+2n\pi,$$

但对于任何整值 n，等式(β)皆成立．根据所得的关于 u 和 v 的式子，我们有

$$Lnz=\ln r+i(\varphi+2n\pi). \tag{25}$$

公式(25)对于所有异于零的复数值 z 定义了函数 Lnz. 这个公式不仅是对于正数，而且也对于负数和复数给出了对数的定义．

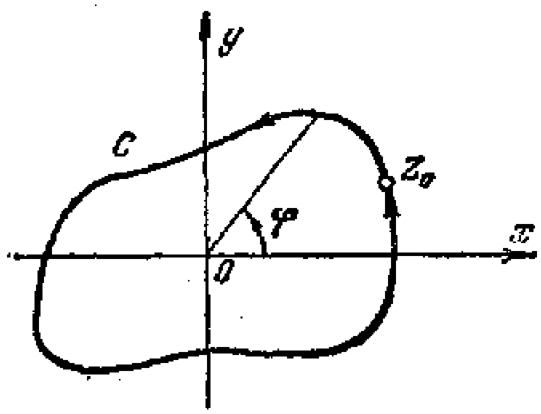


图 4

关于 Lnz 所得到的式子包含有一个任意的整数 n. 这就是说，Lnz 是一个多值函数．对于任意一个 n，我们就得到了函数 Lnz 的一个可能的值．假若把 n 固定，我们就得到这函数的一个可能的值．

但 Lnz 的不同的值相互间是有机地联系着的．事实上，比如

说，我们在 z_0 点固定 $n=0$. 设 c 是一条通过 z_0 点并包围坐标轴原点的闭曲线(图 4). 我们现在假定 z 点沿 c 连续变动. 当 z 移动时，辐角 φ 将连续增加，而当 z 点经过了整个闭曲线时，φ 即增大 2π. 于是，在固定 z_0 点的对数值

$$(Lnz)_0=ln\,r_0+i\varphi_0$$

之后，然后当 z 沿着包围坐标轴原点的闭曲线移动时连续地改变这个函数值，我们在回到 z_0 点时就得到了另外的一个函数值

$$(Lnz)_0=ln\,r_0+i(\varphi_0+2\pi).$$

这证明了函数的任意一个值可以连续地变到另一个值. 为此，就必须让 z 点绕坐标轴原点连续地转过必要的回数. 点 $z=0$ 叫做函数 Lnz 的支点.

假若我们希望只限于讨论函数 Lnz 的一个值，我们就必须禁止 z 点描过包围 $z=0$ 点的闭曲线. 这是可以做到的，只需从坐标轴原点向无限远点引一条连续曲线，然后禁止 z 点穿过这条线即可，这条线叫做割线(或剖线). 假若 z 点在一张具有割线的平面上变动，即已经不可能从 Lnz 的一个值变到另一个值，因而当我们从对数在某一点 z_0 的一个固定的值出发时，我们在每一点只能得到一个对数值. 这样分离出来的函数 Lnz 的值叫做它的一支.

比如说，假如我们沿 Ox 的负的部分引一割线，我们就得到了 Lnz 的一支，它的辐角只限于在范围

$$(2k-1)\pi<\varphi\leqslant(2k+1)\pi$$

内变动，这里的 k 是任一整数.

当我们只考虑对数的一支时，我们可以来研究它的微分. 设

$$r=\sqrt{x^2+y^2},\ \varphi=\operatorname{arc\,tg}\frac{y}{x},$$

容易证明 Lnz 满足柯西–黎曼条件，而它的导数，在根据(比如说)公式(22)计算之后，即等于

$$\frac{dLnz}{dz}=\frac{1}{z}.$$

我们要特别指出，Lnz 的导数已经是一单值函数.

§2. 复变函数与数学物理问题的关系

与流体力学的关系 柯西-黎曼条件把数学物理中的问题与复变函数论联系在一起。我们现在用流体力学中的问题来说明这点。

在流体的各种可能的运动中，**驻定运动**起着特别重要的作用。驻定运动乃是这样的一种运动，对于这种运动，速度在空间的分布状况不因时间的改变而改变。比如说，一个观察者，当他站在桥上注视着流过桥墩的河水时，他就看到河水流动的一个驻定相图。有时，流水对于某些随着某一物体移动的观察者来说，它就变成驻定的。假若在汽船行动时，站在岸上的人们注视着水的波动，那么对于这些观察者，水的运动相图就不是驻定的，但对于汽船上的观察者，水流就是驻定的。对于以定速度飞行的飞机上的乘客来说，因飞机所激动的空气的运动也是驻定的。

在驻定运动之下，通过空间中某一定点的流体质点的速度向量 $\mathbf{V}$ 不因时间而改变。假若对于移动着的观察者来说，运动是驻定的，那么，在随观察者移动着的坐标系中具有固定坐标的那种点，速度向量不因时间而改变。

在流体的各种运动中，**平面平行运动**具有特别重要的意义。平面平行运动是一流动，在这流动之下，质点的速度处处与某一平面平行，而且速度的分布相图在所有与所给平面平行的平面上都是一样的。

假若我们设想流体的质量是无限的，流向垂直于一圆柱体的母线，则在所有与母线垂直的平面上，速度分布的相图都是一样的，因而流体的运动是一个平面平行运动。有时，流体的运动可以近似地看作平面平行运动。比如说，假若我们希望定出气流速度在与机翼垂直的平面上的相图，那么，当这平面不太靠近机身或翼尾时，空气的运动可以近似地看作平面平行运动。

我们现在要来说明，如何采用复变函数论来研究驻定平面平

行运动．这时我们将假定流体是不可压缩的，就是说，它的密度不因压力的改变而改变．比如水就有这种性质，但当我们研究空气的运动时，假若运动的速度不太大，即使空气也可看作是不可压缩的．空气之为不可压缩这一假定，当运动速度不超过音速 ($c=330$ 公尺/秒) 的 0.6—0.8 倍时，并不因此产生值得注意的歪曲．

流体的流动可以由它的质点的速度的分布表示出来．假若流动是平面平行运动，那么只须知道质点在一张与运动进行的方向平行的平面上的速度即可．

我们现在用 $\mathbf{V}(x, y, t)$ 记质点在时间 t 这一瞬间通过坐标为 x, y 这一点的速度．现在我们所讨论的是驻定运动，故 $\mathbf{V}$ 与时间无关．向量 $\mathbf{V}$ 可以看作由它在坐标轴上的射影所定义．我们现在来研究流体质点的轨道．对于驻定运动来说，从空间中所给的点出发的那些质点的轨道不因时间而改变．假若速度场已知，就是说，假若作为 x, y 的函数的速度的分量为已知，则利用质点的速度必常与轨道相切这一点，质点的轨道即可定出．由此即得

$$\frac{dy}{dx}=\frac{v(x, y)}{u(x, y)}.$$

所得出的方程是轨道的微分方程．驻定运动的质点轨道称为流线．在运动平面上的每一点有一条流线通过．

流函数的概念非常重要．我们现在固定任意一条流线 C_0，并考虑一条想象的渠道，这渠道由一柱面（其母线与流体运动的平面平行）范围而成，这柱面过流线 C_0 和另一条流线 C_1 以及两张与运动所在平面平行且相距 1 单位的平面（图 5）．假若来研究我们渠

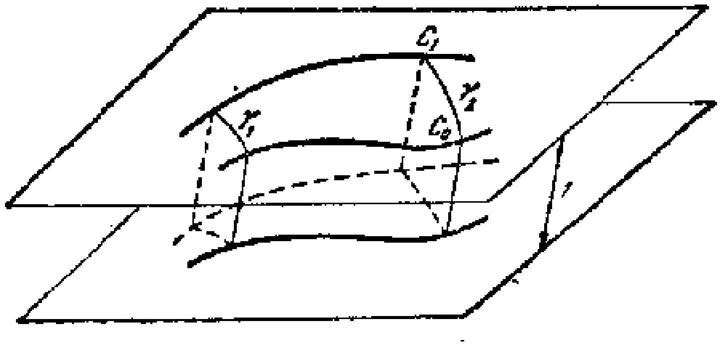


图 5

道的任意两个横截面 γ_1 和 γ_2，则在单位时间内流经截面 γ_1 和 γ_2 的流体的总量相等．事实上，在由侧面 C_1，C_0 和 γ_1，γ_2 所定义的容器内，在密度一定之下，流体的总量不可能发生变化．另一方面，渠道 C_0 和 C_1 的侧壁是由流线做成的，因此，流体在经过它时不会流到别处去，因而在单位时间内，从 γ_1 流进来多少流体，也从 γ_2 流出去多少流体．

一个函数 $\psi(x, y)$，假若它在流线 C_1 上取定值，其值等于流体在单位时间内流经在曲线 C_0 和 C_1 上所作起来的渠道的横截面的总流量，则此函数即称为**流函数**．

流函数除了一个与初始流通曲线 C_0 的选取有关的常数之外，完全决定．若流函数已知，则流线的方程显然是

$$\psi(x, y)=\text{const}.$$

流速的分量可由流函线的导数来表示．要想得出这一表达式，我们现在考虑由过所给的点 $M(x, y)$ 的流线 C，和过邻近点 $M'(x, y+\Delta y)$ 的流线 C'，以及两张相距一单位且与运动平面平行的平面所构成的渠道．我们现在来计算流体在时间 dt 内流经渠道的横截面 MM' 的总量 q．

一方面，由流函数的定义，

$$q=(\psi'-\psi)\,dt.$$

另一方面，q 等于在横截面 MM' 上的每一点所引的向量 $\mathbf{V}dt$ 的全部做成的体积（图 6）．若 MM' 很小，则我们可以假定整个 MM' 上的 $\mathbf{V}$ 为一常量，且等于 M 点的 $\mathbf{V}$ 值．所得的平行六面

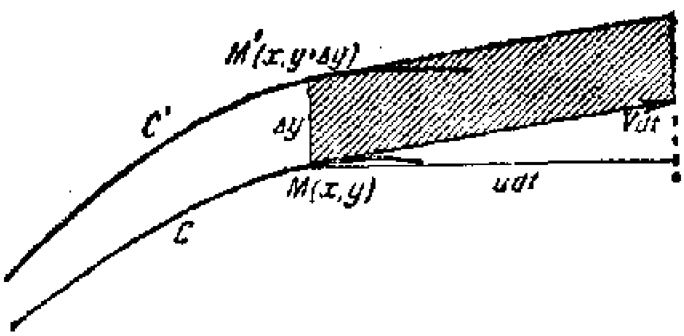


图 6

体的底的面积等于 $\Delta y\times 1$（在图 6 中，一层的单位厚度没有画出），而高则等于向量 $\mathbf{V}dt$ 在 Ox 轴上的射影，即 $u\,dt$，因此，

$$q\approx u\Delta y\,dt,$$

因之，$u\Delta y\approx\Delta\psi$.

以 Δy 除此等式两边，并取极限，则得

$$u=\frac{\partial\psi}{\partial y}. \tag{26}$$

对于第二分量作同样论证即得

$$v=-\frac{\partial\psi}{\partial x}. \tag{26'}$$

要想定出速度场，则除流函数之外，还须引进第二个函数．这函数的引进和流体质点的转动有关．假若我们把流体的各个小质点想像为一刚体，则一般说来，它们都有旋转运动．现若流体的运动是从静止状态发生的，且流体的各个质点之间又没有内部摩擦，则流体中的质点就不可能发生转动．这种没有质点转动的运动叫做**无旋涡运动**，它在研究流体中物体的运动时起着很重要的作用．在流体力学中已经证明，对于无旋涡运动，存在第二个函数 $\varphi(x,y)$，利用这一函数，速度的分量可用函数

$$u=\frac{\partial\varphi}{\partial x},\quad v=\frac{\partial\varphi}{\partial y} \tag{27}$$

表出；函数 φ 叫做流动的**速度位**．我们现在进一步来讨论具有速度位的运动．

试将速度分量关于流函数和关于速度位的公式加以比较，我们即可得出下述重要结论．

不可压缩流体的流动的速度位 $\varphi(x,y)$ 和流函数 $\psi(x,y)$ 满足柯西-黎曼方程

$$\frac{\partial\varphi}{\partial x}=\frac{\partial\psi}{\partial y},$$
$$\frac{\partial\varphi}{\partial y}=-\frac{\partial\psi}{\partial x}. \tag{28}$$

换言之，复变函数

$$w=\varphi(x,\ y)+i\psi(x,\ y)$$

是一个可微分的复变函数． 反之，假若我们从一可微分复变函数的导数出发，则它的实部分和虚部分满足柯西–黎曼方程，因而可以认为是一不可压缩流体的流动的速度位和流函数．函数 w 叫做**流动的特征函数**．

我们现在来讨论 w 的导数的意义．利用（比如说）公式 (22)，则得

$$\frac{dw}{dz}=\frac{\partial\varphi}{\partial x}+i\frac{\partial\psi}{\partial x}.$$

由(27)和(26′)，即得

$$\frac{dw}{dz}=u-iv,$$

或取共轭数，有

$$u+iv=\overline{\left(\frac{dw}{dz}\right)},\tag{29}$$

在这里，$\frac{dw}{dz}$ 上面的横线是表示须取与之共轭的数．

于是，流动的速度向量等于流动的特征函数的导数的共轭值．

流体的平面平行流动举例 我们现在来讨论几个例子．令

$$w=Az,\tag{30}$$

在这里，A 为一复数．由(29)，即得

$$u+iv=\overline{A}.$$

因此，线性函数 (30) 就定义了一个具有常速度向量的流体运动．若令

$$A=u_0-iv_0,$$

则将 w 的实部分和虚部分分开，即得

$$\varphi(x,\ y)=u_0x+v_0y,$$

$$\psi(x,\ y)=u_0y-v_0x,$$

于是，与速度向量平行的直线是流线(图 7)．

作为第二个例子，我们现在来讨论函数

$$w=Az^2,$$

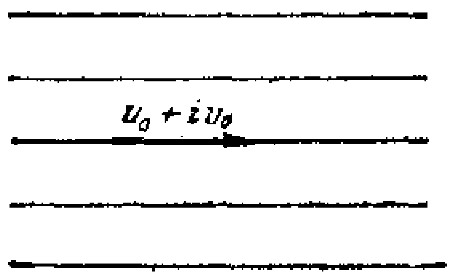


图 7

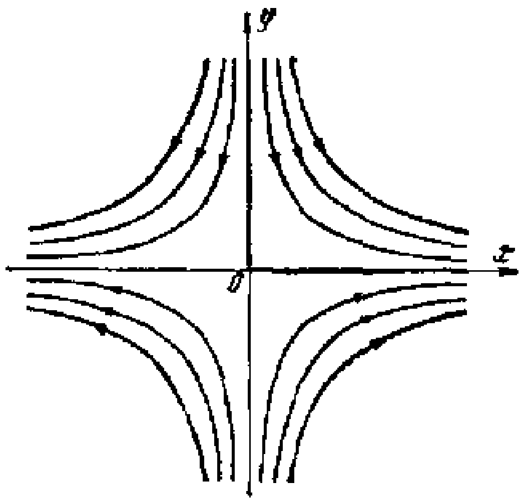


图 8

在这里，A 是一实的常数．为了要画出流动的相图，我们现在来定出流线．在现在的情形下，有

$$\psi(x,\ y)=2Axy,$$

而流线的方程则为

$$xy=\text{const}.$$

这是一双曲线，它以坐标轴为其渐近线（图 8）．箭头所指的是当 $A>0$ 时，质点沿流线运动的方向．Ox 轴和 Oy 轴也是流线．

若在流体中摩擦非常小，则在将任意一条流线代之以一固定的壁之后，并不影响其余的流动．流体的质点将沿所置的壁滑动．利用这一原理，并将该壁沿坐标轴放置（在图 8 中，这是用粗线画出），则在所讨论的例子中，我们就得出了流体流经角点的无旋涡流动的相图．

函数

$$w=a\left(z+\frac{R^2}{z}\right) \tag{31}$$

给出了一个重要的流动，在这里，a 和 R 是正实数．

流函数为

$$\psi=a\left(y-\frac{R^2y}{x^2+y^2}\right),$$

因而流线的方程为

$$y-\frac{R^2y}{x^2+y^2}=\mathrm{const}.$$

特别，若取常数为零，则或得 $y=0$，或得 $x^2+y^2=R^2$；因而半径为 R 的圆周是一流通曲线。若我们置一固定的物体于此流通曲线的内部，则得一围绕圆柱的流动。这流动的流通曲线的相图如图 9 所示。流动速度可由公式(29)

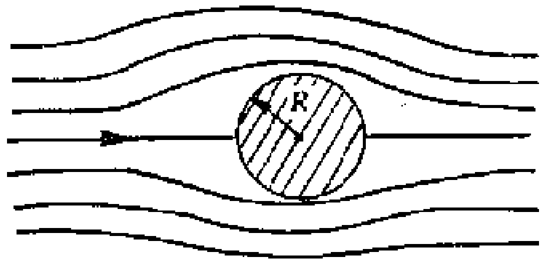


图 9

$$u+iv=a\left(1-\frac{R^2}{z^2}\right)$$

定出。

在离圆柱很远之处，我们有

$$\lim_{z\to\infty}(u+iv)=a,$$

即在离圆柱很远之处，速度趋于定量，因而流动变为等速的。于是，公式(29)决定了流体的一个流动，它流经圆柱，且在远处为等速流动。

机翼理论的基本概念．茹可夫斯基定理 将复变函数论运用来研究流体的平面平行流动，就使得茹可夫斯基和恰普里金在空气动力学中得出了一个重要的发现．在研究流体流经一物体时，引起他们发现了机翼支持力的形成规律．要想对于导致这一发现的概念的形成给予一些观念，我们还需要来讨论一个关于流体运动的具体例子．我们现来考虑特征函数

$$w=\frac{\Gamma}{2\pi i}\,Lnz.$$

在这里，Γ 是一实的常数．虽然函数 w 不是单值函数，但它的导数

$$\frac{dw}{dz}=\frac{\Gamma}{2\pi i}\frac{1}{z} \tag{32}$$

却是单值的，因而我们的函数就一意地决定了某一流体运动的速度场。令 $z=re^{i\theta}$，则速度位和流函数可以利用公式(25)算出：

$$\varphi=\frac{\Gamma}{2\pi}\theta,\ \psi=-\frac{\Gamma}{2\pi}Ln\,r.$$

上式中第二式指出，流通曲线是圆 $r=\text{const}$ (图 10).

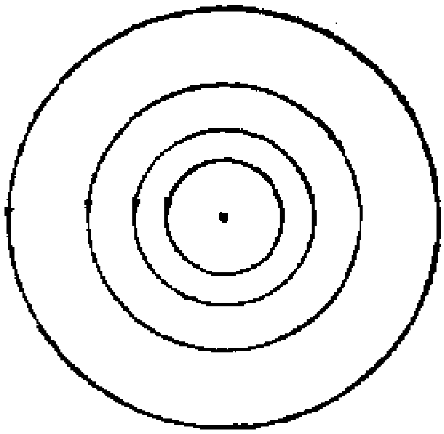


图 10

流动速度是由公式(29)定出

$$u+iv=-\frac{\Gamma}{2\pi i}\ \frac{1}{z}.$$

特别，由此可以推出，速度向量的大小为

$$V=|u+iv|=\frac{|\Gamma|}{2\pi}\ \frac{1}{r},$$

即在流线上的每一点，速度为常数. 作更详细一些的研究，可知当 $\Gamma>0$ 时，流动系反时针进行，当 $\Gamma<0$ 时，则顺时针进行.

若我们将某一条流线代之以一固定的边界，则即得流体绕圆柱的一环状运动. 这样的运动叫做**环流**.

我们的运动的位已经不再是一单值函数. 在沿环绕圆柱的一闭曲线绕一周时，位即改变一个量 Γ. 位的这一改变叫做**流动的环流量**.

假若我们将流经圆柱的流动的特征函数 (31) 加上 (顺时针绕动的)环流的特征函数，我们就得到了一个新的特征函数

$$w=a\left(z+\frac{R^2}{z}\right)-\frac{\Gamma}{2\pi i}\ Lnz. \tag{33}$$

这特征函数也表示一个环绕半径为 R 的圆柱的流动. 事实上，在半径为 R 的圆周上，流函数为一常数，因为在这圆周上，两个被加项的虚部分都是常数. 由函数 (33) 所定义的流动的速度当 $z\to\infty$ 时仍趋于 a. 这说明了，对于任何的 Γ 值，特征函数(33)决定了一个以前进运动绕过圆柱的流动. 在图 11 中，我们画出了这一流动当 $\Gamma>0$ 时的形势. 这一流动已经不是对称的了，流体正朝着圆柱流进来和流出去的点 a 和 b 已经向下移了. 我们所讨论的流动的位是一个非单值函数. 它在绕圆柱转一周时的改变

为 Γ.

当流体流经圆柱时，根据对称性，通常出现的是由函数(32)所定义的对称流动；但对于其他非对称物体，一般则出现具有非单值位的流动．下面我们将要来阐明这项事实的物理根源．利用复变函数论方法，我们可以定出环绕任何一种形状的物体的各种可能的流动．在下一节中，我们将要谈到这些方法．无论是环绕圆柱或环绕任何别的物体，我们都可造出具有单值位和非单值位的流动．

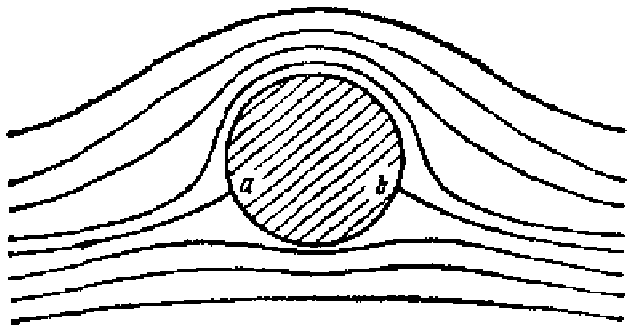
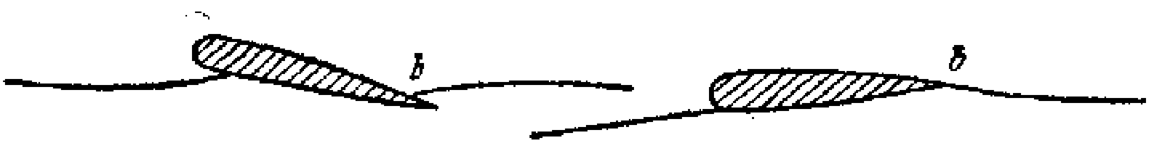


图 11

当研究气体流经机翼时，我们将要来研究后面有一尖的边缘的物体．机翼的横断面常常后面呈尖状．假若对于这种横断面我们造出了一具有单值位的流动，则气流从横断面上离开之点结果并不与尖点一致(图 12,(a))．这样的流动在物理上是不可能的．(在这样的流动之下，在横断面的尖端就会出现无穷大的速度和无限稀薄.)只有在流动中使得 b 点和机翼的尖点一致的那种流动(图 12,(b))才是唯一可能的流动，这种流动一般具有非单值位，即为一环流．这种流动的环流量 Γ 仍然是被定义为当循一条包围机翼的闭曲线绕一周时位的改变量.



(a) (b)

图 12

下面的假说叫做恰普里金假说．我们可以作出气流的一流动，它环绕机翼的横断面流过，且气流从横断面后部的边缘离开.

茹可夫斯基的重要发现是在于：当在气流中有环流量出现时，

就会发生一股上举之力作用于机翼，其方向与冲击过来的气流的速度 a 垂直，其量等于

$$\rho a\Gamma,$$

在这里，ρ 是流体的密度，Γ 是环流量（图 13）.



图 13

这一假说构成了茹可夫斯基关于使机翼上举的力的定理，它是一切近代空气动力学的基础. 茹可夫斯基定理的证明不在我们这篇短文范围之内，我们现在仅指出，对于这定理现今所采用的证明乃是依据复变函数的积分理论.

在茹可夫斯基和恰普里金的工作中所奠定的空气动力学基础已在许多苏联学者的工作中得到了发展.

在其他数学物理问题上的应用 复变函数论在流体力学中不仅是在机翼理论方面，而且还在许多别的问题上也得到了广泛的应用.

但函数论的应用范围并不限于流体力学，它很广阔. 函数论方法在许许多多数学物理问题中曾被大量用到. 为要说明这点，我们现回到柯西–黎曼条件

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y},$$

$$\frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x},$$

并由此导出一个为复变解析函数的实部所满足的方程. 若将上面两方程中前一个关于 x 微分，将第二个关于 y 微分，然后相加，则得

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0.$$

这方程(我们曾于第六章中遇到过)叫做**拉普拉斯方程**. 物理和力学中有大量的问题和拉普拉斯方程有关. 比如说，若在某一物体上出现了热的平衡，则温度即满足拉普拉斯方程. 重力场或静电场

的研究也与这一方程有关．在研究流体通过多孔介质的渗透时，我们也得到了拉普拉斯方程．在所有这些与解拉普拉斯方程有关的问题中，函数论方法都有着广泛的应用．

不仅拉普拉斯方程的研究，而且比之更为一般的数学物理方程的研究也多半与复变函数论有关．这类问题的最重要的例子当中，有一个就是弹性理论的平面问题，复变函数论在这些领域中的应用基础是由苏联学者科洛索夫和穆斯海里什维里所奠定的．

§3．复变函数与几何的关系

可微分函数的几何性质 恰如在研究实变函数的时候一样，在复变函数论中，函数的几何解释具有很重要的地位．可以大胆地说，复变函数的几何性质不仅可以用来直观的表示出函数的解析性质，而且还给这种理论带来了一个专门的课题．与函数的几何理论有关的这部分问题叫做**函数的几何理论**．正如前面所说的，从几何观点看来，复变函数 $w=f(z)$ 是 z 平面到 w 平面的一个映照．这一映照也可以由两个实变函数

$$u=u(x,\ y),$$
$$v=v(x,\ y)$$

给出．假若我们希望在某一点的非常小的邻近研究映照的特性，我们可以将这函数展成泰勒级数，并只展到展开式的主项

$$u-u_0=\left(\frac{\partial u}{\partial x}\right)_0(x-x_0)+\left(\frac{\partial u}{\partial y}\right)_0(y-y_0)+\cdots,$$
$$v-v_0=\left(\frac{\partial v}{\partial x}\right)_0(x-x_0)+\left(\frac{\partial v}{\partial y}\right)_0(y-y_0)+\cdots,$$

这里的导数是在 $(x_0,\ y_0)$ 点取的．因之，在这点的附近，所有的映照可以大致看作是一个仿射[1]映照

$$u-u_0=a(x-x_0)+b(y-y_0),$$
$$v-v_0=c(x-x_0)+d(y-y_0),$$

1) 参看第三章(第一卷)§11．

此时，

$$a=\left(\frac{\partial u}{\partial x}\right)_0,\quad b=\left(\frac{\partial u}{\partial y}\right)_0,$$

$$c=\left(\frac{\partial v}{\partial x}\right)_0,\quad d=\left(\frac{\partial v}{\partial y}\right)_0.$$

我们现在来研究由一个解析函数在某一点 $z=x+iy$ 的附近所做出的映照的性质。设 C 是从 z 点引出的一条曲线；在 w 平面上，与这曲线相应的点做成一条从 w 点引出的曲线 Γ. 假若 z' 点是 z 点的一邻近点，w' 是与之相应的点，则当 $z'\to z$ 时，$w'\to w$，而且

$$\frac{w'-w}{z'-z}\to f'(z). \tag{34}$$

特别，由此可以得出，

$$\left|\frac{w'-w}{z'-z}\right|\to|f'(z)|. \tag{35}$$

这可以用语言陈述如下。

设 C 为在 z 平面上从给定的 z 点引出的任一曲线，Γ 为 w 平面上与 C 相应的曲线。则 C 上过 z 点的弦与 Γ 上相应的弦之比的极限，无论 C 如何，皆为一定，或者，有如人们所说，w 平面上的线素与 z 平面上在所有点的线素之比与从 z 点所引的曲线无关.

量 $|f'(z)|$，它可以用来说明 z 点的线素的大小，叫做 z 点的**伸缩系数**.

我们现在假定在某一点 z 导数，$f'(z)\neq0$，于是，量 $f'(z)$ 就有一个完全确定的辐角[1]. 我们现在利用 (34) 来计算它，

$$\arg\frac{w'-w}{z'-z}=\arg(w'-w)-\arg(z'-z),$$

但 $\arg(w'-w)$ 是弦 ww' 与实轴的交角 β'，$\arg(z'-z)$ 是弦 zz' 与实轴的交角 α'. 假若我们分别用 α 和 β 来记曲线 C 和 Γ 在 z 点和 w 点的切线与实轴的交角(图 14)，则当 $z'\to z$ 时，$\alpha'\to\alpha$，$\beta'\to\beta$，因此，取极限，即得

$$\arg f'(z)=\beta-\alpha. \tag{36}$$

1) 参看第四章(第一卷)§3.

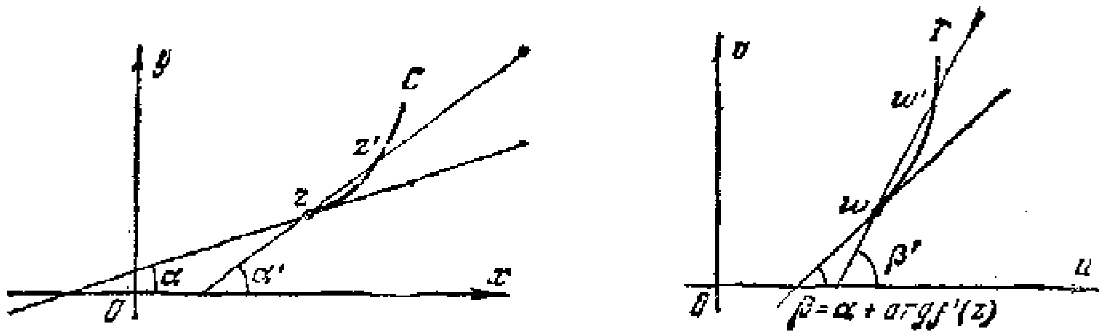


图 14

这等式指出，$\arg f'(z)$ 等于一个角 φ，假若要从曲线 C 在 z 点的切线的方向转到曲线 Γ 在 w 点的切线的方向，就必须转过这一个角．由于这种性质，**$\arg f'(z)$ 叫做 z 点的映照的回转角．**

读者容易从等式(36)推出下面的命题．

当从 z 平面变到 w 平面时，任意一条过所有点的曲线的切线皆转过同一个角．

假若 C_1 和 C_2 是从 z 点引出的两条曲线，Γ_1 和 Γ_2 是从 w 点引出的相应曲线，则 Γ_1 和 Γ_2 在 w 所构成的角等于 C_1 和 C_2 在 z 点所构成的角．

于是，假若在一点 z，$f'(z)\neq 0$，则在解析函数 $f(z)$ 在该点做成的映照之下，所有线素皆按同一比例伸缩，而相应方向之间的交角则不变．

具有所述性质的映照称为**保角映照．**

从上面所证明的映照的几何性质，假若在某一点 z_0，有 $f'(z_0)\neq 0$，则在这点的附近，我们自然就会期望，在 z_0 点的某一很小的邻域之内，映照是相互一意的，就是说，不仅每一点 z 只有一点 w 与之对应，而且反过来：每一点 w 也只有一点 z 与之对应．这实际上是可以严格证明的．

为了完全的想象出保角映照与其它各种映照是怎样的不同，把任意一种映照在某一点的一个很小的邻域内来加以研究将会是有好处的．设函数 u, v 是做出映照的两个函数，假若我们来研究一下，u, v 的泰勒展开式的主项，我们就得到

$$u-u_0=\left(\frac{\partial u}{\partial x}\right)_0(x-x_0)+\left(\frac{\partial u}{\partial y}\right)_0(y-y_0)+\cdots,$$
$$v-v_0=\left(\frac{\partial v}{\partial x}\right)_0(x-x_0)+\left(\frac{\partial v}{\partial y}\right)_0(y-y_0)+\cdots.$$

假若在(x_0, y_0)点的很小的邻域之内将高阶项略去不算，则我们的映照的变化状况就和一个仿射变换一样．假若这变换的行列式不为零，

$$\Delta=\left(\frac{\partial u}{\partial x}\right)_0\left(\frac{\partial v}{\partial y}\right)_0-\left(\frac{\partial u}{\partial y}\right)_0\left(\frac{\partial v}{\partial x}\right)_0\neq 0,$$

则这变换是可逆的．假若$\Delta=0$，则在表示这变换在(x_0, y_0)点附近的变化情况时，就必须考虑高阶项[1]．

当$u+iv$是一解析函数时，我们可以利用柯西-黎曼条件将关于y的导数用关于x的导数来表示，于是即得

$$\Delta=\left(\frac{\partial u}{\partial x}\right)_0^2+\left(\frac{\partial v}{\partial x}\right)_0^2=\left|\left(\frac{\partial u}{\partial x}\right)_0+i\left(\frac{\partial v}{\partial x}\right)_0\right|^2=|f'(z_0)|^2,$$

即当$f'(z_0)\neq 0$时，变换为可逆的．若令$f'(z_0)=r(\cos\varphi+i\sin\varphi)$，则

$$\left(\frac{\partial u}{\partial x}\right)_0=\left(\frac{\partial v}{\partial y}\right)_0=r\cos\varphi,$$
$$\left(\frac{\partial u}{\partial y}\right)_0=-\left(\frac{\partial v}{\partial x}\right)_0=-r\sin\varphi,$$

而在(x_0, y_0)点附近的映照的形式则为

$$u-u_0=r[(x-x_0)\cos\varphi-(y-y_0)\sin\varphi]+\cdots,$$
$$v-v_0=r[(x-x_0)\sin\varphi+(y-y_0)\cos\varphi]+\cdots,$$

这公式指出，若函数$w=u+iv$为解析，则在(x_0, y_0)点附近的映照可以归纳成：转一个φ角，并依系数r伸缩．事实上，从解析几何中可以知道，括号内的式子是平面旋转一个φ角的旋转公式，而用r乘一乘，则表示伸缩r倍．

为了想象出，在使得$f'(z)=0$的那种点，映照可能出现一些什么，我们把函数

---

1) 在后一种情形下，即当$\Delta=0$时，该变化已经不叫仿射变换．关于仿射变换，可参考第三章(第一卷)§11．

$$w=z^n \tag{37}$$

加以研究将会是有好处的．这函数的导数 $w'=nz^{n-1}$ 当 $z=0$ 时为零．映照(37)最好是利用极坐标或复数的三角表示来加以研究．令

$$z=r(\cos\varphi+i\sin\varphi),$$

$$w=\rho(\cos\theta+i\sin\theta).$$

注意在复数相乘时，模相乘而辐角相加，则得

$$z^n=r^n(\cos n\varphi+i\sin n\varphi),$$

因而有

$$\rho=r^n,\quad \theta=n\varphi.$$

从后面一个式子可以看出，z 平面上的射线 $\varphi=\text{const}$（常数）变成 w 平面上的射线 $\theta=n\varphi=\text{const}$．因之，$z$ 平面上两条射线之间大小为 α 的交角变成大小为 $\beta=n\alpha$ 的交角．z 平面到 w 平面上的映照已经不再是一对一的．事实上，若已经给定了一个以 ρ 为模，以 θ 为辐角的点 w，则它可以作为 n 个以 $r=\sqrt[n]{\rho}$ 为模，以

$$\varphi=\frac{\theta}{n},\quad \frac{\theta}{n}+\frac{2\pi}{n},\ \cdots,\ \frac{\theta}{n}+\frac{2\pi}{n}(n-1)$$

为辐角的点的象而得出．当自乘 n 次幂时，相应点的模即等于 ρ，辐角即等于

$$\theta,\ \theta+2\pi,\ \cdots,\ \theta+2\pi(n-1),$$

因为把辐角加上一个 2π 倍数的量并不改变该点的几何位置，故 w 平面上所有的象皆为同一．

保角映照 假若解析函数 $w=f(z)$ 把 z 平面上的一个域 D 一对一地变到了 w 平面上的域 Δ，我们就说它做出了一个从域 D 到域 Δ 上的保角映照．

保角映照在函数论及其应用中的作用可以从下面一个差不多是一目了然的定理确定下来．

假若 $\zeta=F(w)$ 是域 Δ 内的一个解析函数，则复合函数 $F[f(z)]$ 是域 D 内的一个解析函数．这定理可从以下等式推出：

$$\frac{\Delta\zeta}{\Delta z}=\frac{\Delta\zeta}{\Delta w}\cdot\frac{\Delta w}{\Delta z}.$$

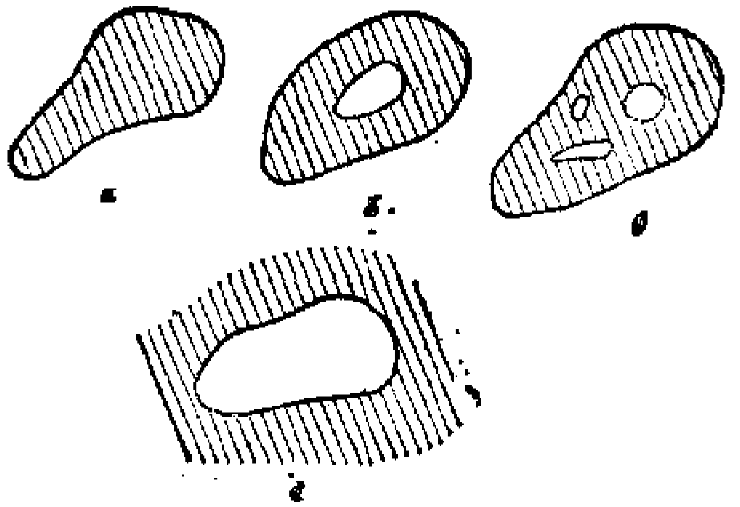


图 15

注意一下函数 $\zeta=F(w)$ 和 $w=f(z)$ 为解析，我们就可断定，右边的两个因子皆有极限，因而在域 D 内的每一点，比值 $\frac{\Delta\zeta}{\Delta z}$ 具有唯一的极限 $\frac{d\zeta}{dz}$. 这证明了函数 $\zeta=F[f(z)]$ 为解析.

所证明的定理指出了，在域 Δ 中研究解析函数可以化为在域 D 中去研究解析函数. 假若域 D 的几何构造比较简单，那么函数的研究就可因此化简.

所有那些我们必须要在其中来进行解析函数的研究的域之中，最重要的一类要算是单连通域类. 所谓单连通域是指它的周界由一条边所组成的那种域(图 15, (a))，以别于周界分解成几条边的那种域(例如图 15, (b) 和 15, (c) 中所画的那种域).

我们要注意，有时我们也需要在处于曲线外面，而不是在处于曲线里面的那种域中来研究函数，假若这种域的周界只由一条边所组成，这种域还是叫做单连通域(图 15, (d)).

下面一个著名的黎曼定理是保角映照理论的基础.

对于任何单连通域 Δ，我们皆可造出一个解析函数. 它把以原点为心，以 1 为半径的圆保角地映照到域 Δ 上去，并使得圆心映照到域 Δ 的一个预先给定的点 w_0，且把过圆心的任意一个预先给定的方向变到过 w_0 点的一个任意预先给定的方向. 这定理指出

了，在任何单连通域内的复变函数的研究可以化归（例如）在单位圆内定义的函数的研究．

我们现在大致来解释一下如何把上面所说的事实用到机翼理论的问题上去．现在假定我们希望要去研究在一个具有某种形式的横断面的机翼附近的气流．

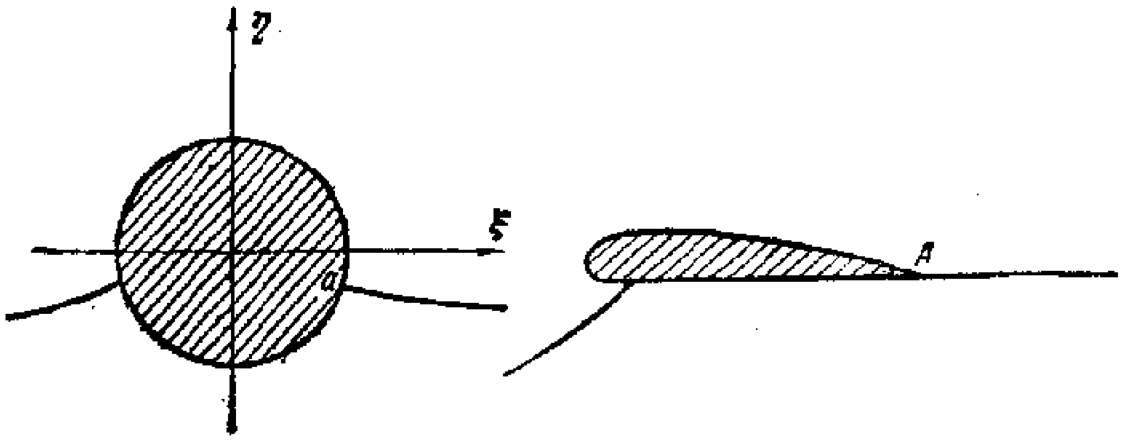


图 16

假若我们已经会把横断面外面的流场保角地映照到圆外面的域上去，那么我们就可以把那已经造好了的圆外附近气流的特征函数的表达式利用来造出横断面附近气流的特征函数．

设 ζ 是圆所在的平面，z 是横断面所在的平面，$\zeta=f(z)$ 是一个函数，它把横断面外面的域映照到圆外面的域上去，而且有

$$\lim_{z\to\infty}\zeta=\infty.$$

以 a 记圆上面与横断面的尖点 A 相应的点．我们现在作一个环流，它绕圆周流过，并在 a 点有一流动点（图 16）．这函数我们以 $W(\zeta)$ 记之，

$$W(\zeta)=\Phi+i\Psi.$$

这气流的流动曲线系由方程

$$\Psi=\text{const}$$

所定义．我们现来研究函数

$$w(z)=W[f(z)],$$

并令 $w=\varphi+i\psi$.

我们现在证明，$w(z)$ 是在横断面周围流经 A 点的气流的特征函数。我们首先证明由函数 $w(z)$ 所定义的气流流过横断面。要想证明这点，我们须先证明横断面的周线是一流动曲线，即在横断面的周线上，有

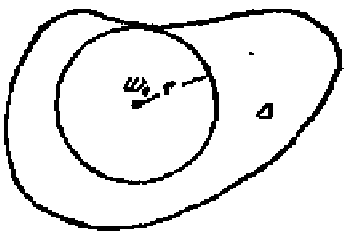


图 17

$$\psi(x, y) = \text{const}.$$

但由此可知，

$$\psi(x, y) = \psi(\xi, \eta),$$

且横断面上的点 (x, y) 与圆周上的点 (ξ, η) 相对应，在圆周上则有 $\psi(\xi, \eta) = \text{const}$.

同理，我们可以简单地证明 A 是一流动点。我们可以证明，假若适当地选取圆上的一道流的流动速度，则在横断面的周围，我们可以得到一气流，它在横断面上的一道流具有任何预先给定的速度。

保角映照在复变函数论和它的应用之中有着重要的作用。由此就引出了一个问题，要寻求一个保角映照，把一个域映照到一个具有给定的几何形状的域上去。在许多最简单然而有用的情形，这一问题可利用初等复变函数来解决。但在一般情形下，初等函数却无济于事。正如我们已经说过的，黎曼曾经说出了保角映照理论的一般性定理，但他并未对这定理作出严格的证明。在好几十年的时间里经过若干大数学家的努力，才对黎曼定理得出了一个完善的证明。

有一些方法密切结合着黎曼定理的各种证明方法发展了起来，利用这些方法我们可以近似地造出域的保角映照。真正要作出一个保角映照把一个域映照到另外一个域上去，这有时是一件非常困难的事情。对于研究函数的若干一般性的性质，通常并不需要真正知道把一个域映照到另一个域上去的保角映照，而只需知道这种映照的某些几何性质即可。这就使得我们要广泛地去研究保角映照的几何性质。为了关于这类定理给读者们一个观念，我们现在试举一例加以说明。

设 z 平面上以坐标轴原点为心，以 1 为半径的圆已经映照到某一域 Δ 上去，而且原点变到域 Δ 内的 w_0 点（图 17）．假若我们现在来研究任意一个把圆映照到域 Δ 上去的映照，则对于这映照在 $z=0$ 点的变化情形，我们不可能作任何的结论．对于保角映照，我们就有下述的著名命题．

在坐标轴原点的伸缩量不超过以 w_0 点为心内切于域 Δ 的周界的圆的半径的 4 倍：

$$|f'(0)|\leqslant 4r.$$

苏联数学家们有许多工作是从事于保角映照理论中许多问题的研究的．在这些工作里，对于许多值得注意的保角映照类得出了精确的公式，研究了保角映照的近似计算方法，并得出了一系列关于保角映照的一般性几何定理．

拟保角映照　保角映照与解析函数的研究，或与满足柯西-黎曼条件

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y},$$

$$\frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}$$

的一对函数的研究密切有关．在若干数学物理问题中，我们会遇到比较普遍的一类微分方程组．这类方程组与某些映照有关，这些映照在平面 Oxy 上的每一点的附近皆具有某种几何性质，且把一平面映照到另一平面．为了说明这点，我们现在来研究下面一组微分方程

$$\frac{\partial u}{\partial x}=p(x,\ y)\frac{\partial v}{\partial y},\qquad (38)$$

$$\frac{\partial v}{\partial x}=-p(x,\ y)\frac{\partial u}{\partial y}.$$

若 $p(x,\ y)=1$，则这组微分方程即退化为柯西-黎曼方程．在一般情形下，假若 $p(x,\ y)$ 是任意的函数，我们也可以把方程组(38)的每一组解释成是 Oxy 平面到 Ouv 平面上的一个映照．现在来研究我们的映照在 $(x_0,\ y_0)$ 点附近的几何性质．为此，假定 $(x_0,\ y_0)$

点的邻域很小，我们可以只保留函数 u 和 v 关于 $x-x_0$ 和 $y-y_0$ 的展开式的第一项，并把映照看成仿射变换

$$u-u_0=\left(\frac{\partial u}{\partial x}\right)_0(x-x_0)+\left(\frac{\partial u}{\partial y}\right)_0(y-y_0),$$
$$v-v_0=\left(\frac{\partial v}{\partial x}\right)_0(x-x_0)+\left(\frac{\partial v}{\partial y}\right)_0(y-y_0). \tag{39}$$

若函数 u, v 满足方程组 (38)，则对于这仿射变换，就出现了下面的性质.

中心在 (x_0, y_0) 点，主轴与坐标轴平行，且半轴之比为

$$\frac{b}{a}=p(x_0, y_0)$$

的椭圆在 Ouv 平面上即变为以 (u_0, v_0) 为心的圆.

我们现在来证明这一命题. Ouv 平面上以 (u_0, v_0) 点为心的圆的方程为

$$(u-u_0)^2+(v-v_0)^2=\rho^2.$$

以 $u-u_0$ 和 $v-v_0$ 的由 x 和 y 表出的表达式代入，则得 Oxy 平面上相应曲线的方程

$$\left[\left(\frac{\partial u}{\partial x}\right)_0^2+\left(\frac{\partial v}{\partial x}\right)_0^2\right](x-x_0)^2$$
$$+2\left[\left(\frac{\partial u}{\partial x}\right)_0\left(\frac{\partial u}{\partial y}\right)_0+\left(\frac{\partial v}{\partial x}\right)_0\left(\frac{\partial v}{\partial y}\right)_0\right](x-x_0)(y-y_0)$$
$$+\left[\left(\frac{\partial u}{\partial y}\right)_0^2+\left(\frac{\partial v}{\partial y}\right)_0^2\right](y-y_0)^2=\rho^2.$$

现在利用方程组 (38) 将函数 v 的导数用函数 u 的导数表出. 由此即得

$$\left[\left(\frac{\partial u}{\partial x}\right)_0^2+p^2\left(\frac{\partial u}{\partial y}\right)_0^2\right](x-x_0)^2$$
$$+\frac{1}{p^2}\left[\left(\frac{\partial u}{\partial x}\right)_0^2+p^2\left(\frac{\partial u}{\partial y}\right)_0^2\right](y-y_0)^2=\rho^2.$$

若令
$$a=\frac{\rho}{\sqrt{\left(\frac{\partial u}{\partial x}\right)_0^2+p^2\left(\frac{\partial u}{\partial y}\right)_0^2}},$$

$$b=\frac{p\rho}{\sqrt{\left(\frac{\partial u}{\partial x}\right)_0^2+p^2\left(\frac{\partial u}{\partial y}\right)_0^2}},$$

则上面的方程即化为

$$\frac{(x-x_0)^2}{a^2}+\frac{(y-y_0)^2}{b^2}=1.$$

于是，变为圆的曲线实际上是一具有上述性质的椭圆.

假若我们不是去讨论由展开式的第一项得出的仿射变换，而是讨论精确的变换，则椭圆的半轴越小，刚才对于映照所得出的性质就越接近真实的情况.

我们可以这样说，这性质对于无穷小椭圆成立.

于是，从方程(38)可以推知，变为圆的无穷小椭圆的半轴的方向和半轴之比在每一点皆可定出. 实际上，这一几何性质已完全描绘出微分方程组(38)的特性，就是说，假若函数 u 和 v 所作出的映照具有所述的几何性质，则它们就满足这一组方程. 因此，研究方程(38)的解的问题和研究具有所述性质映照的问题是等价的.

特别，我们要指出，对于柯西-黎曼方程，这一性质可叙述如下：

以 $(x_0,\ y_0)$ 点为心的无穷小圆变为以 $(u_0,\ v_0)$ 点为心的无穷小圆.

很广泛的一类数学物理方程的研究可以归结为具有下述几何性质的映照的研究.

对于变数平面上的每一点 $(x,\ y)$，我们给定了两个椭圆的半轴之比和半轴的方向. 要求作出 Oxy 平面到 Ouv 平面上的映照，使得第一族中的无穷小椭圆变为第二族中以点 $(u,\ v)$ 为心的无穷小椭圆.

对于与这种一般的方程组有关的映照，首先是由苏联数学家拉夫伦捷夫研究起来的. 这些映照名为拟保角映照. 研究由微分方程组所定义的映照这种想法，使得我们有可能把解析函数论中的方法推广到非常广泛的一类问题上去. 拉夫伦捷夫和他的学生

们就曾进行拟保角映照的研究；而且这些研究在数学物理、力学、几何中的若干问题上得到许许多多的应用. 值得提及的是：拟保角映照的研究在解析函数论本身也是很有前途的.

自然，在这里我们不可能详细谈论复变函数论中的几何方法在各方面的应用.

§4. 线积分。柯西公式及其推论

复变函数的积分 对于解析函数性质的研究，复变函数的积分这一概念起着最为重要的作用. 复变函数沿曲线求积分这一概念相当于实变函数的定积分这一概念. 我们现在来研究平面上从 z_0 开始到 z 终止的一条弧 C 和一函数 $f(z)$，这函数在包含弧 C 的一个域内有定义. 我们将弧 C 用点(图 18)

$$z_0,\ z_1,\ \cdots,\ z_n = z$$

分成一些小的部分，然后来考虑和数

$$S = \sum_{k=1}^{n} f(z_k)(z_k - z_{k-1}).$$

假若函数 $f(z)$ 为连续，而弧 C 又为有限长，则如对于实函数一样，我们可以类似地证明，在将分点的数目 n 增大，使得相邻两个分点之间的距离趋于零时，和数 S 有一完全确定的极限. 这极限叫做沿弧 C 的积分，以

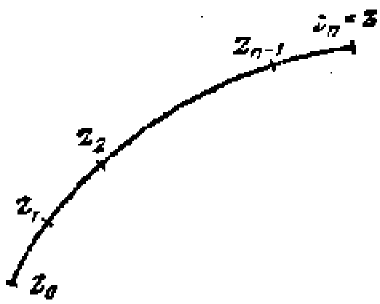


图 18

$$\int_C f(z)\,dz$$

记之. 我们要注意，在定义积分时，我们已经选定了弧 C 的始点和终点，换言之，我们已经选定了确定的沿曲线 C 运动的方向.

我们容易证明下面一切简单的积分性质.

1° 两函数之和的积分等于被加函数的积分和，

$$\int_C [f(z)+g(z)]\,dz = \int_C f(z)\,dz + \int_C g(z)\,dz.$$

2° 常数因子可以拿到积分符号之外，

$$\int_C Af(z)dz = A\int_C f(z)dz.$$

3° 若弧 C 是弧 C_1 和 C_2 之和，则

$$\int_C f(z)dz = \int_{C_1} f(z)dz + \int_{C_2} f(z)dz.$$

4° 若 $\bar{C}$ 是弧 C 按反方向所取的弧，则

$$\int_{\bar{C}} f(z)dz = -\int_C f(z)dz.$$

所有这些性质，对于和数 S 来说皆很显然，而对于积分来说，则可从取极限得出.

5° 若弧 C 之长等于 L，而且在弧 C 上有不等式

$$|f(z)| \leqslant M,$$

则

$$\left|\int_C f(z)dz\right| \leqslant ML.$$

我们来证明这一性质. 我们只需对于积分和 S 来证明这一不等式即可，因为这样一来，它对于取极限之后仍然成立，因而对于积分也成立，对于积分和

$$|S| = \left|\sum f(z_k)(z_k - z_{k-1})\right| \leqslant \sum |f(z_k)|\,|z_k - z_{k-1}| \leqslant M\sum |z_k - z_{k-1}|.$$

但位于第二因子的和数等于内接于弧 C 且以 z_k 为顶点的折线各段之长的和. 显而易见，折线之长不会大于曲线之长，因而有

$$|S| \leqslant ML.$$

我们现在来看最简单的函数 $f(z)=1$ 的积分. 显而易见，在这种情形下，有

$$S = (z_1 - z_0) + (z_2 - z_1) + \cdots + (z_n - z_{n-1}) = z_n - z_0 = z - z_0.$$

这证明了 $\int_C 1.dz = z - z_0$.

所得到的结果指出，对于函数 $f(z)=1$，积分的值对于所有连接 z_0 点和 z 点的弧皆相同. 换言之，积分的值仅与积分路径的始

点和终点有关．但容易证明，这一性质并不是对于任何复变函数皆成立．比如说，假若 $f(z)=x$，则经过简单的计算即可指出，

$$\int_{C_1} x\,dz=\frac{x^2}{2}+iyx,\quad \int_{C_2} x\,dz=\frac{x^2}{2},\quad z=x+iy.$$

这里的 C_1 和 C_2 是图 19 中画出的积分路径．

我们把这两个等式的证明留给读者．

解析函数论的重要事实就是下面的柯西定理．

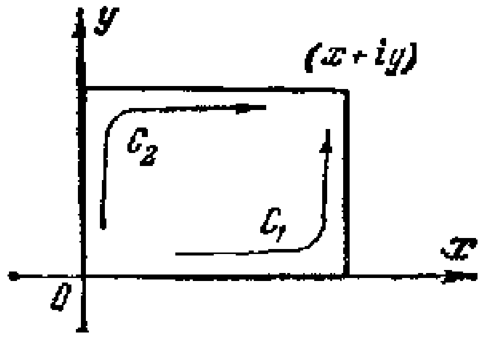


图 19

若 $f(z)$ 在单连通域 D 内的每一点皆可微分，则沿域 D 内任何连接 D 内任意两点 z_0 和 z 的弧所取的积分皆相同．

在这里，我们将不证明柯西定理，对此有兴趣的读者可参考任何一本关于复变函数论的书．我们在这里只叙述这定理的一些重要推论．

首先，柯西定理使得我们可以引入解析函数的不定积分．事实上，我们现在将 z_0 点固定，而来讨论沿一条连接 z_0 和 z 的曲线所取的积分，

$$F(z)=\int_{z_0}^{z} f(\zeta)\,d\zeta.$$

这时我们可以沿任何一条连接 z_0 和 z 的曲线取积分，因为积分的值不因曲线的改变而改变，因而只与 z 点有关．函数 $F(z)$ 叫做 $f(z)$ 的**不定积分**．

$f(z)$ 的不定积分具有导数，它等于 $f(z)$．

在若干的应用上，有几个别的与柯西定理等价的定理比较合用．

假若 $f(z)$ 在一单连通域内处处可以微分，则沿这域内的任何一条闭曲线所取的积分等于零，

$$\int_{\Gamma} f(z)\,dz=0.$$

这是显而易见的，因为闭曲线的始点和终点相同，因而 z_0 和 z 可

以用一条零径连接起来.

以后，凡是提到闭曲线，我们皆指的是依反时针方向所引的一条围道. 假若围道是依时针方向所引，我们将以 $\overline{\Gamma}$ 记之.

柯西积分 上面所说的事实使得我们可以推出下面的柯西基本公式，这公式把可微分函数在闭围道内点的值用函数在围道本身上面的值表示出来

$$f(z)=\frac{1}{2\pi i}\int_C\frac{f(\zeta)}{\zeta-z}\,d\zeta.$$

我们现在证明这一公式. 设 z 固定，ζ 为一自变数. 函数

$$\varphi(\zeta)=\frac{f(\zeta)}{\zeta-z}$$

在域 D 内部，除了点 $\zeta=z$ 之外（在这点分母为零），在每一点 ζ 皆为连续且可微分，这一事实使得我们不能把柯西定理运用到函数 $\varphi(\zeta)$ 和围道 C 上来.

我们现在来考虑以 z 点为圆心以 ρ 为半径的圆周 K_ρ，并证明

$$\int_C\varphi(\zeta)\,d\zeta=\int_{K_\rho}\varphi(\zeta)\,d\zeta. \tag{40}$$

为了这一目的，我们现在做一条辅助的闭围道 Γ_ρ，它是由围道 C，连接 C 和圆周的弧段 γ_ρ，以及依反方向所引的圆周 $\overline{K}_\rho$ 所组成的（图 20）. 围道 Γ_ρ 有如箭头指出. 因为点 $\zeta=z$ 在 Γ_ρ 之外，所以在 Γ_ρ 之内函数 $\varphi(\zeta)$ 处处可微分，因而有

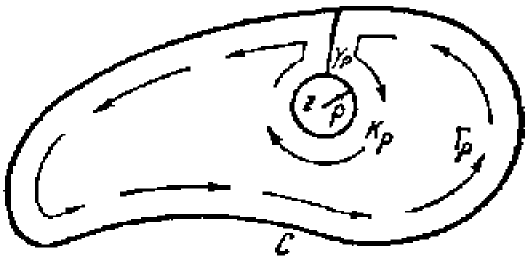


图 20

$$\int_{\Gamma_\rho}\varphi(\zeta)\,d\zeta=0. \tag{41}$$

但围道 Γ_ρ 可以分成四个部分：C，γ_ρ，$\overline{K}_\rho$ 和 $\overline{\gamma}_\rho$，故据积分的性质 3°，我们有

$$\int_{\Gamma_\rho}\varphi(\zeta)\,d\zeta=\int_C\varphi(\zeta)\,d\zeta+\int_{\gamma_\rho}\varphi(\zeta)\,d\zeta$$
$$+\int_{\overline{K}_\rho}\varphi(\zeta)\,d\zeta+\int_{\overline{\gamma}_\rho}\varphi(\zeta)\,d\zeta=0.$$

以沿 K_ρ 及 γ_ρ 所取之积分代替沿 $\overline{K}_\rho$ 及 $\overline{\gamma}_\rho$ 所取的积分，并利用性质 4°，即得

$$\int_{\Gamma_\rho}\varphi(\zeta)d\zeta=\int_C\varphi(\zeta)d\zeta-\int_{K_\rho}\varphi(\zeta)d\zeta=0,$$

这证明了公式(40)。

要想算出(40)的右边，我们现在令

$$\begin{aligned}\int_{K_\rho}\varphi(\zeta)d\zeta&=\int_{K_\rho}\frac{f(\zeta)}{\zeta-z}d\zeta\\&=\int_{K_\rho}\frac{f(\zeta)-f(z)}{\zeta-z}d\zeta+\int_{K_\rho}\frac{f(z)d\zeta}{\zeta-z}\\&=\int_{K_\rho}\frac{f(\zeta)-f(z)}{\zeta-z}d\zeta+f(z)\int_{K_\rho}\frac{d\zeta}{\zeta-z}.\end{aligned}\tag{42}$$

我们先计算第二项．在圆周 K_ρ 上，

$$\zeta=z+\rho(\cos\theta+i\sin\theta).$$

试注意 z 和 ρ 为常数，我们即得

$$d\zeta=\rho(-\sin\theta+i\cos\theta)d\theta=i\rho(\cos\theta+i\sin\theta)d\theta,$$

此外

$$\zeta-z=\rho(\cos\theta+i\sin\theta),$$

故得

$$\int_{K_\rho}\frac{d\zeta}{\zeta-z}=\int_{K_\rho}id\theta=2\pi i.$$

因为绕圆周一周，θ 之全部变动等于 2π．根据(40)和(42)，

$$\int_C\frac{f(\zeta)d\zeta}{\zeta-z}=2\pi if(z)+\int_{K_\rho}\frac{f(\zeta)-f(z)}{\zeta-z}d\zeta.$$

在所得出的等式中，我们现在令 $\rho\to0$ 取极限．此时，左边和右边之第一项保持不变．我们现在证明，第二项的极限等于零．于是，在 $\rho\to0$ 时，我们的等式即变为柯西公式．要想证明第二项当 $\rho\to0$ 时趋于零，我们可注意

$$\lim_{\zeta\to z}\frac{f(\zeta)-f(z)}{\zeta-z}=f'(\zeta),$$

即在积分号下的式子具有有限极限，因而为有界，

$$\left|\frac{f(\zeta)-f(z)}{\zeta-z}\right|<M.$$

运用积分性质 5°，即得

$$\left|\int_{K_\rho}\frac{f(\zeta)-f(z)}{\zeta-z}d\zeta\right|\leqslant M\,2\pi\rho\rightarrow 0.$$

于是，柯西公式的证明即告完成．柯西公式乃是复变函数论的研究中的一个基本工具．

将可微分函数展成幂级数 我们现在利用柯西定理来证明可微分复变函数的两个性质．

任何复变函数，若它在域 D 内具有一阶导数，则它也具有各阶导数．

事实上，在闭曲线内，我们的函数可用柯西积分表出，

$$f(z)=\frac{1}{2\pi i}\int_C\frac{f(\zeta)}{\zeta-z}d\zeta.$$

在积分号下的是 z 的可微分函数；因之，在积分号下取微分，则得

$$f'(z)=\frac{1}{2\pi i}\int_C\frac{f(\zeta)}{(\zeta-z)^2}d\zeta.$$

在积分号下的重新又是一个可微分函数，因此，我们可以再取微分，于是即得

$$f''(z)=\frac{1\cdot 2}{2\pi i}\int_C\frac{f(\zeta)d\zeta}{(\zeta-z)^3}.$$

继续施行微分，我们即得到一般性公式

$$f^{(n)}(z)=\frac{n!}{2\pi i}\int_C\frac{f(\zeta)d\zeta}{(\zeta-z)^{n+1}}.$$

于是，我们就可以计算任何阶导数．要想使得这一证明十分严格，还须证明在积分号下求微分是合法的．我们将不讨论这一部分证明．

第二基本性质是：

若 $f(z)$ 在以 a 点为心的圆 K 内处处可微分，则 $f(z)$ 可展成一在 K 内收敛的台勒级数．

$$f(z)=f(a)+\frac{f'(a)}{1!}(z-a)+\cdots+\frac{f^{(n)}(a)}{n!}(z-a)^n+\cdots.$$

在§1中，我们曾定义复变数的解析函数是一个可展成幂级数的函数．上述的定理说道：任意可微分的复变函数皆是解析函数．这是复变函数的一个特有的性质，在实域中并无与之类似者．

具有一阶导数的实变函数可以在任何一点皆没有二阶导数.

我们现在来证明上述的定理.

设 $f(z)$ 在以 a 点为心的圆 K 内及其周界上皆具有导数. 则在 K 内, 函数 $f(z)$ 可用柯西积分表出

$$f(z)=\frac{1}{2\pi i}\int_C \frac{f(\zeta)\,d\zeta}{\zeta-z}. \tag{43}$$

记 $\zeta-z=(\zeta-a)-(z-a)$, 则得

$$\frac{1}{\zeta-z}=\frac{1}{(\zeta-a)-(z-a)}=\frac{1}{\zeta-a}\,\frac{1}{1-\dfrac{z-a}{\zeta-a}}. \tag{44}$$

试注意点 z 系在圆内, 而 ζ 则在圆周上, 故得

$$\left|\frac{z-a}{\zeta-a}\right|<1,$$

因之, 根据几何级数公式, 有

$$\frac{1}{1-\dfrac{z-a}{\zeta-a}}=1+\left(\frac{z-a}{\zeta-a}\right)+\cdots+\left(\frac{z-a}{\zeta-a}\right)^n+\cdots, \tag{45}$$

而右边的级数为收敛. 利用 (44) 及 (45), 我们可以把公式 (43) 表成

$$f(z)=\frac{1}{2\pi i}\int_C\left[\frac{f(\zeta)}{\zeta-a}+(z-a)\frac{f(\zeta)}{(\zeta-a)^2}+\cdots\right.$$
$$\left.+(z-a)^n\frac{f(\zeta)}{(\zeta-a)^{n+1}}+\cdots\right]d\zeta.$$

我们现在将括弧内之级数施行逐项积分. 这样做之所以合法, 可以严格地加以论证. 于是, 在各项内将与 ζ 无关的因子 $(z-a)^n$ 提出括号之后, 我们就得到

$$f(z)=\frac{1}{2\pi i}\int_C\frac{f(\zeta)\,d\zeta}{\zeta-a}+\frac{z-a}{2\pi i}\int_C\frac{f(\zeta)\,d\zeta}{(\zeta-a)^2}+\cdots$$
$$+\frac{(z-a)^n}{2\pi i}\int_C\frac{f(\zeta)\,d\zeta}{(\zeta-a)^{n+1}}+\cdots.$$

现利用逐次导数的积分公式, 我们可以记

$$\frac{1}{2\pi i}\int_C \frac{f(\zeta)d\zeta}{(\zeta-a)^{n+1}}=\frac{f^{(n)}(a)}{n!},$$

于是即得

$$f(z)=f(a)+\frac{f'(a)}{1!}(z-a)+\cdots+\frac{f^{(n)}(a)}{n!}(z-a)^n+\cdots.$$

我们已经证明了可微分复变函数可展成幂级数．反之，可表成幂级数的函数是可微分的．它的导数可以从逐项微分级数得到（这一运算之为合法可以严格证明）．

整函数 幂级数只在某一圆内给出了函数的解析表示．这圆的半径等于从圆心到与之最为接近的使得函数不再为解析的点（函数的奇点）的距离．

在解析函数中，自然也有一类函数，它们对于变数的任何有限值皆为解析．这种函数可以用一个幂级数来表示，这幂级数对于变数 z 的一切值皆为收敛，这种函数名为 z 的整函数．假若我们是讨论在坐标轴原点的展开式，则整函数可以表成形如

$$G(z)=C_0+C_1z+C_2z^2+\cdots+C_nz^n+\cdots$$

的级数．若在这级数中，所有的系数，从某一个开始，皆为零，则这函数就是一个多项式，或整有理函数

$$P(z)=C_0+C_1z+\cdots+C_nz^n.$$

假若在展开式中有无限多项异于零，则此整函数叫做超越的．下面是几个这类函数的例子：

$$e^z=1+\frac{z}{1!}+\frac{z^2}{2!}+\cdots,$$

$$\sin z=\frac{z}{1!}-\frac{z^3}{3!}+\frac{z^5}{5!}-\cdots,$$

$$\cos z=1-\frac{z^2}{2!}+\frac{z^4}{4!}-\cdots.$$

在研究多项式时，方程

$$P(z)=0$$

的根的分布问题起着重要的作用，或者，更普遍一些，我们可以提出关于使得多项式取所与的值 A

$$P(z)=A$$

的点的分布问题.

高等代数学中的基本定理说：任何多项式至少在一个点取所与的值 A. 这一性质已经不可能搬到任意的整函数上来. 例如函数 $w=e^z$ 在 z 平面上任何一点皆不为零. 但我们却有下列的皮卡尔定理：任何整函数取任意的值无限多次，至多除一个值之外.

关于使得整函数取所与的 A 值的那种点在平面上的分布问题乃是整函数论中的一个中心问题.

多项式的根的数目等于它的次数. 多项式的次数与 $|P(z)|$ 当 $|z|\to\infty$ 时的增长速度密切有关. 事实上，我们可以写

$$|P(z)|=|z|^n\cdot\left|a_n+\frac{a_{n-1}}{z}+\cdots+\frac{a_0}{z^n}\right|,$$

因为当 $z\to\infty$ 时，第二因子趋于 $|a_n|$，故当 $|z|$ 甚大时，n 次多项式的增大与 $|a_n||z|^n$ 相似. 于是，我们可以看出，n 越大，$|P_n(z)|$ 当 $|z|\to\infty$ 时增长越快，多项式的根越多. 看起来，似乎对于整函数来说这项规律仍然继续有效. 但对于整函数 $f(z)$ 来说，根的数目一般皆是无限，因此，根数的问题就没有什么意义. 但我们还是可以讨论方程

$$f(z)=a$$

在半径为 r 的圆内根的数目 $n(r, a)$，并研究当 r 增大时，这一数目如何变化. $n(r, a)$ 增长的速度与整函数在半径 r 的圆内的最大模 $M(r)$ 的增长速度密切有关. 正如已经说过的，对于整函数来说，可以存在一个例外的值 a，使得方程对它没有一个根. 对于所有另外的值 a，数目 $n(r, a)$ 增长的速度可以和量 $\ln M(r)$ 增长的速度相比. 在这里，我们不可能对这些规律性作更精确的叙述.

整函数的根的分布特性和数论中的问题有关，它使得我们对于黎曼捷塔函数 $\zeta(s)$ 找出了许多重要的性质[1)]，根据这些性质，我们曾经证明了许多有关素数的定理.

分式函数或半纯函数 整函数类可以看作是代数多项式类的

1) 参考专门讨论数论的第十章.

一推广. 从多项式出发，我们可以得出比之更为宽广的有理函数类

$$R(z)=\frac{P(z)}{Q(z)},$$

这是两个多项式之比.

同理，从整函数出发，我们自然可以做出新的函数类，由两个整函数 $G_1(z)$ 和 $G_2(z)$ 之比作成的函数

$$f(z)=\frac{G_1(z)}{G_2(z)},$$

叫做**分式函数**或**半纯函数**. 这样得出的函数类在数学分析中起着重大的作用. 在初等复变函数中，像

$$\operatorname{tg} z=\frac{\sin z}{\cos z},\quad \operatorname{ctg} z=\frac{\cos z}{\sin z}$$

等就属于半纯函数类.

半纯函数已经不是在全复平面上皆为解析. 在使得分母 $G_2(z)$ 为零的那些点，函数 $f(z)$ 变为无限. $G_2(z)$ 的根在复平面上做成一孤立点的集. 在这些点的附近，函数 $f(z)$ 当然不可能展成泰勒级数，但在这种点 a 的附近，半纯函数却可以表成一个含有 $(z-a)$ 的若干个负方次的幂级数，

$$f(z)=\frac{C-m}{(z-a)^m}+\cdots+\frac{C-1}{z-a}+C_0$$
$$+C_1(z-a)+\cdots+C_n(z-a)^n+\cdots. \qquad (46)$$

当 z 点接近 a 点时，$f(z)$ 的值趋于无限. 使得一解析函数在该点变为无限的孤立奇点称为该函数的**极点**. 在 a 点函数的解析性质的损失系由展开式(46)中具有 $z-a$ 的负方次的诸项所约制. 或

$$\frac{C_{-m}}{(z-a)^m}+\cdots+\frac{C_{-1}}{(z-a)}$$

描述出半纯函数在接近奇点时的变化情形，称为展开式 (46) 的**主要部分**. 半纯函数在奇点附近变化情形的特点由它的主要部分所决定. 在许多情形下，当我们知道一半纯函数在它所有奇点的附近的展开式的主要部分之后，就可把这函数造出来. 比如说，假若

函数 $f(z)$ 为有理函数，且在无限远点为零，则此函数即等于它在所有极点附近的展开式的主要部分之和，而对于有理函数来说，极点的数目则是有限的，

$$f(z)=\sum_{(k)}\left[\frac{C_{-m_k}^{(k)}}{(z-a_k)^{m_k}}+\cdots+\frac{C_{-1}^{(k)}}{z-a_k}\right].$$

在一般情形下，有理函数可以表成它的各主要部分以及某一多项式之和，

$$f(z)=\sum_{(k)}\left[\frac{C_{-m_k}^{(k)}}{(z-a_k)^{m_k}}+\cdots+\frac{C_{-1}^{(k)}}{z-a_k}\right]$$
$$+C_0+C_1z+\cdots+C_mz^m. \tag{47}$$

(47)式给出了有理函数的表示式，在这式子中，奇点在函数的构造上所起的作用表现得很清楚．在有理函数的各种各样的应用之中，(47)式这样的表示式用起来很为方便，而且在说明函数的奇点如何决定函数整个的结构时，它也大有益处．看起来，所有的半纯函数也正如有理函数一样，可以就它在各极点的主要部分来构造．我们现在引述(例如)函数 $\operatorname{ctg} z$ 的详细的表示式，但未加证明．函数 $\operatorname{ctg} z$ 的极点可以作为方程

$$\sin z=0$$

的根而得出，即位于点 $\cdots,\ -k\pi,\ \cdots,\ -\pi,\ 0,\ \pi,\ \cdots,\ k\pi,\ \cdots$．我们可以证明，函数在极点 $z=k\pi$ 的幂级数展开式的主要部分为

$$\frac{1}{z-k\pi},$$

而函数 $\operatorname{ctg} z$ 即等于各极点的主要部分的和

$$\operatorname{ctg} z=\frac{1}{z}+\sum_{k=1}^{\infty}\left[\frac{1}{z-k\pi}+\frac{1}{z+k\pi}\right]. \tag{48}$$

半纯函数按主要部分展成级数的展开式之所以重要，是因为在这种表达式中，函数的奇点已经清楚地表出，而且这样的解析表示还使我们可以算出在全平面上使得函数具有定义的诸点的函数值．

在研究许许多多分析中重要的函数类时，半纯函数论是非常重要的．特别，我们必须强调一下它在数学物理方程理论中的意

义．积分方程论可以使得我们回答数学物理方程理论中许多重要的问题．而积分方程论的建立，则大有赖于半纯函数论中的基本命题．

直到现在，泛函分析中与数学物理最为密切有关的一部分——运算子理论——的发展仍常有赖于解析函数论中的一些事实．

论函数的解析表示 上面我们已经看到，在函数可以微分的任何点的附近，该函数都可利用幂级数来定义．对于整函数来说，幂级数在全平面上收敛，因而在所有使得函数有定义的点给出了函数的解析表示．假若函数不是整函数，则如我们所知，函数的泰勒级数只在某一圆内收敛，这圆的圆周经过与圆心最为接近的函数的奇点．由是，这幂级数就不可能在所有使得函数有定义的点皆能算出函数的值．因而这解析函数就不可能在它的整个定义域之内用一个幂级数来定义．对于半纯函数来说，按主要部分展开的展开式就是它的解析表示，它在函数的整个定义域之内定义了函数．

假若函数不是整函数，但它在某一圆之内定义，或者，假若我们有一个函数，它在某一域之内定义，但我们只希望在某一圆内来研究它，那么，在表示它的时候，我们就可以使用泰勒级数．当我们在一个不是圆形的域中来研究函数时，我们就遇到一个问题，是否可以找到函数的一个解析表示，它适于在整个所说的域中表示函数．在一个圆内表示一解析函数的幂级数，它以最简单的多项式 a_nz^n 作为它的项．自然会发生这样的问题，能不能在一任意给定的域中将函数展开成多项式级数．当级数的每一项又可能利用算术运算来计算的时候，我们又得到了一个重新从最简单的算术运算出发来表示函数的工具．我们所提出的问题的一般答案可由下面的定理得出：

在任意一个边界系由一条曲线做成的域内定义的解析函数，可以展成多项式级数

$$f(z)=P_0(z)+P_1(z)+\cdots+P_n(z)+\cdots.$$

上述的定理只一般地回答了函数在一任意给定的域内可否展开成多项式级数这一问题．但这一定理还不可能使我们根据给定的函数造出级数，如同对于泰勒级数所出现的那样．这定理只是更迅速地提出了如何将函数展开成多项式级数这一问题，而不是解决了这一问题．关于根据所给的多项式或它的某些性质造出多项式级数的问题，造出收敛最快的级数以及与该函数变化情形的特点密切相关的级数的问题，根据已经给定的用来表示函数的多项式级数以研究函数的构造的问题，等等，就构成了广阔发展的用多项式逼近函数的理论．关于这种理论的建立，苏联科学家们起着非常重要的作用，他们在这方面得出了一系列带根本性的结果．

§5. 唯一性和解析拓展

解析函数的唯一性 解析函数的最重要性质之一就是它的唯一性．

假若在域 D 内给定了两个解析函数，它们在某一条属于域内的曲线 C 上相同，则它们在整个域内全同．

这定理的证明很简单．设 $f_1(z)$ 和 $f_2(z)$ 是域 D 内的两个解析函数，它们在曲线 C 上一致．差

$$\varphi(z)=f_1(z)-f_2(z)$$

是域 D 内的解析函数，且在 C 上为零．我们现证明，在域 D 内任何一点皆有 $\varphi(z)=0$．事实上，假若在域 D 内存在一点 z_0（图 21），于此，$\varphi(z_0)\neq 0$，则我们可将曲线 C 延长到点 z_0，并沿所得到的曲线 Γ 走向 z_0，直至函数不等于零为止．假若 ζ 是在 Γ 上用这种方法所得到的最后一点．若 $\varphi(z_0)\neq 0$，则 $\zeta\neq z_0$，且在曲线 Γ 的紧接 ζ 点之后的任何线段上不恒为零．我们现在来证



图 21

明这是不可能的．事实上，在曲线 Γ 的紧接 ζ 点之前的部分 Γ_ζ 上，我们有 $\varphi(z)=0$．我们可以只用到函数 $\varphi(z)$ 在 Γ_ζ 上的值就

能算出 $\varphi(z)$ 在 Γ_ζ 上的所有导数，因而 $\varphi(z)$ 的所有导数皆为零. 特别，在 ζ 点，有

$$\varphi(\zeta)=\varphi'(\zeta)=\cdots=\varphi^{(n)}(\zeta)=\cdots=0.$$

我们现在将函数 $\varphi(\zeta)$ 在 ζ 点展成泰勒级数. 展开式的所有系数皆为零，于是，在某一以 ζ 点为心且属于域 D 内的圆内有

$$\varphi(z)=0.$$

特别，由此可以推出，等式 $\varphi(z)=0$ 在曲线 Γ 的紧接 ζ 点之后的某一线段上继续成立. 由假定 $\varphi(z_0)\neq 0$ 即导致一矛盾.

所证明的定理指出，若已知解析函数在某一曲线段或域的某一部分的值，则这些值即唯一地决定了函数在所给定的域内各处的值. 于是，函数在变数平面上各个部分之间的值互相密切有关.

要想懂得解析函数唯一性的意义，我们必须回忆一下，依照复变函数的一般定义，在变数值和函数值之间可以容许任何的对应规律. 在这样的定义之下，当然谈不到一个地区的函数值就唯一决定了平面上其它地区的函数值. 我们现在看到，复变函数的可微分性这一唯一的要求显得是这样的有力，以致决定了函数在各个地区的值之间的关系.

我们还须特别指出，在实变函数范围内，函数的可微分性并不引出类似的推论. 事实上，我们可以造出几个函数，它们可微分任意多次，在 Ox 轴的某一区间内相同，而在其它的点则不相等. 例如当 x 取负值时等于零的函数可以这样定义，使得它当 x 取正值时异于零，且有任意次的连续导数. 比如说，当 $x>0$ 时，令

$$f(x)=e^{-\frac{1}{x}},$$

即可做到这点.

解析拓展与全解析函数 在定义复变函数时，函数的定义域常为定义函数的方法本身所限制. 我们现在来讨论一个完全初等的例子. 设函数由级数

$$f(z)=1+z+z^2+\cdots+z^n+\cdots \tag{49}$$

所定义. 我们都知道，这级数在以坐标轴原点为心的单位圆内收

敛，且在这圆外发散．因此，由(49)式所定义的解析函数只在这圆内定义．另一方面，我们知道，级数在圆 $|z|<1$ 内的和可以表成公式

$$f(z)=\frac{1}{1-z}. \tag{50}$$

(50)式对于任何值 $z\neq 1$ 都有意义．根据唯一性定理，(50)式给出了唯一在圆 $|z|<1$ 内与级数(49)之和相等的解析函数．于是，原先只在单位圆内定义的函数，我们已经把它拓展到全平面．

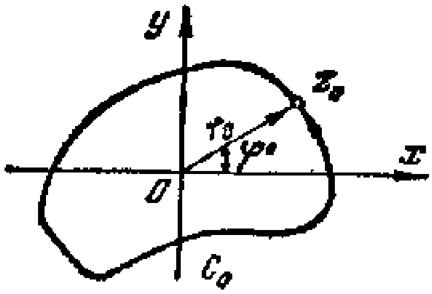


图 22

假若我们有一个函数 $f(z)$，它在某一域 D 内有定义．又若存在另外的一个函数 $F(z)$，它在一个包含 D 的域 Δ 内有定义，且在 D 内与 $f(z)$ 相等，则根据唯一性定理，$F(z)$ 在 Δ 内的值是唯一定义的．

函数 $F(z)$ 称为 $f(z)$ 的**解析拓展**．一个解析函数，假若它除非失去它的解析性质，否则就不能拓展到这域的外面去，这样的函数我们称之为**完全的**．例如在全平面上定义的整函数就是完全函数．半纯函数也是完全函数，它除了它的极点之外处处有定义．但也存在解析函数，它的完全定义域是一个有界域．我们现在不去谈论这些相当复杂的例子．

完全解析函数这一概念使得我们有必要去研究非单值的复变函数．我们现在以函数

$$Lnz=lnr+i\varphi$$

为例来说明这点，在这里，$r=|z|$，$\varphi=\arg z$．假若在 z 平面上的某一点 $z_0=r_0(\cos\varphi_0+i\sin\varphi_0)$ 去研究函数的某一初值

$$(Lnz)_0=lnr_0+i\varphi_0,$$

那么，当 z 点沿某一曲线 C 移动时，我们的解析函数就可连续不断的拓展下去．正如上面所说，我们容易看出，假若 C_0 是一条从 z_0 点出发，包围坐标轴原点然后重新回到 z_0 点的闭曲线（图 22），则当 z 点描过 C_0 时，我们在 z_0 点仍回到初值 lnr_0，但角 φ 则增加

2π. 这证明了，当我们沿曲线 C 连续不断的拓展函数 Lnz 时，绕曲线 C 一次，函数即增加 $2\pi i$. 假若 z 点描过这闭曲线 n 次，则代替初值

$$(Lnz)_0 = lnr_0 + i\varphi_0,$$

我们得到了新的值

$$(Lnz)_n = lnr_0 + (2\pi n + \varphi_0)i.$$

假若 z 点按负方向描过曲线 C m 次，我们即得

$$(Lnz)_{-m} = lnr_0 + (-2\pi m + \varphi_0)i.$$

上述的论证说明了，在复变数平面上，我们不可避免地要去研究 Lnz 的互相之间有着联系的各个值. 函数 Lnz 是无限多值的. 对于多值函数 Lnz，点 $z=0$ 处境很为奇特，绕它一次，我们就可从一个值变到另一个值. 容易证明，假若 z 点描过一条不环绕坐标轴原点的闭曲线，则此时 Lnz 的值不变. 点 $z=0$ 叫做函数 Lnz 的支点.

一般来说，假若对于某一函数 $f(z)$，当绕过 a 点一周之后，我们即从它的一个值变到另一个值，则 a 点即称为函数 $f(z)$ 的支点.

我们现在讨论另一个例子. 令

$$w = \sqrt[n]{z}.$$

如同上面一样，可以指出，这函数也是多值的，且取 n 个值

$$\sqrt[n]{r}\left(\cos\frac{\varphi}{n} + i\sin\frac{\varphi}{n}\right),\quad \sqrt[n]{r}\left(\cos\frac{\varphi+2\pi}{n} + i\sin\frac{\varphi+2\pi}{n}\right),$$

$$\cdots,\quad \sqrt[n]{r}\left(\cos\frac{\varphi+2\pi(n-1)}{n} + i\sin\frac{\varphi+2\pi(n-1)}{n}\right).$$

假若我们从函数的一个值

$$w_0 = \sqrt[n]{r_0}\left(\cos\frac{\varphi_0}{n} + i\sin\frac{\varphi_0}{n}\right)$$

出发，并绕一条包围坐标轴原点的闭曲线移动，我们就可得出函数所有不同的值，因为每绕原点一周，角 φ 即增加 2π.

绕闭曲线移动 $(n-1)$ 次，则从 $\sqrt[n]{z}$ 的第一个值出发，我们就得到了它的 $(n-1)$ 个其余的值. 绕曲线移动 n 次，则即得根值

$$\sqrt[n]{z_0}=\sqrt[n]{r_0}\left(\cos\frac{\varphi_0+2n\pi}{n}+i\sin\frac{\varphi_0+2\pi n}{n}\right)$$
$$=\sqrt[n]{r_0}\left(\cos\frac{\varphi_0}{n}+i\sin\frac{\varphi_0}{n}\right),$$
即我们又回到出发时用的根值.

多值函数的黎曼面 有一个非常直观的几何表示方法，可表出多值函数的性质.

我们再来讨论函数 Lnz, 并在 z 平面上沿 Ox 轴的正的部分引一割线. 假若 z 点不准通过这割线，则我们就不可能从 Lnz 的一个值连续地变到另一个值. 若从 z_0 点拓展 Lnz 则我们只能得到 Lnz 的一个值.

像这样在割开了的 z 平面上得到的单值函数，叫做函数 Lnz 的一个单值支. Lnz 的所有的值就分配在无限多个单值支上，

$$\ln r+i\varphi,\ 2\pi n<\varphi\leqslant 2\pi(n+1).$$

容易证明，第 n 支在割线的下面部分正好取第 $(n+1)$ 支在割线的上面部分的值.

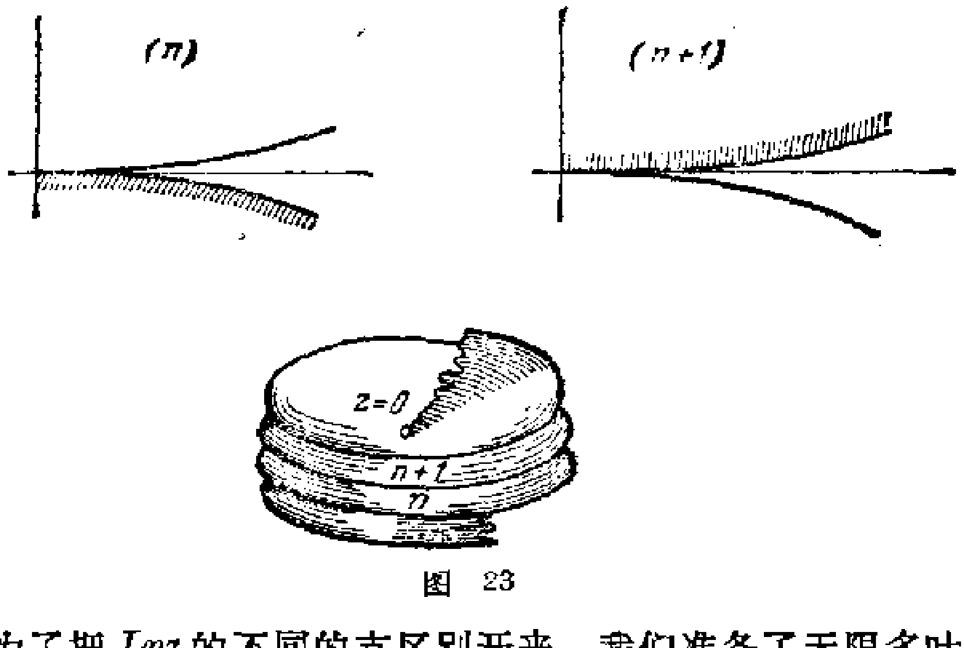


图 23

为了把 Lnz 的不同的支区别开来，我们准备了无限多叶同样的 z 平面，它们都沿 Ox 轴的正部分割开. 我们把与函数的第 n 支相对应的 z 值表示在第 n 叶 z 平面上. 属于不同的 z 平面上然而具有同样坐标的点这时对应于同一个数 $x+iy$; 只有复数在第 n 叶

上的像才标志着我们是在研究对数的第 n 支.

为了在几何上表示出对数的第 n 支在第 n 叶平面的割线的下面部分的值与对数的第 $(n+1)$ 支在第 $(n+1)$ 叶平面的割线的上面部分的值相同, 我们把第 n 叶平面的割线的下面部分与第 $(n+1)$ 叶平面的割线的上面部分接合起来, 于是就把这两叶平面粘在一起. 利用这种做法, 我们就得到一张多叶面, 它的结构和螺梯相似(图 23). 这时 $z=0$ 这一点就起着螺梯中柱的作用.

假若点从一叶变到另一叶, 则复数即回到它出发时的值, 而函数 Lnz 即从这一支变到另一支.

刚才所造出来的面叫做函数 Lnz 的**黎曼面**. 黎曼第一个提出了构造这种面的概念, 用它来表现出多值解析函数的特征, 并指出这一概念的内容丰富.

我们现在还来作函数 $w=\sqrt{z}$ 的黎曼面. 这函数是一个二值函数, 并以坐标轴原点为其支点.

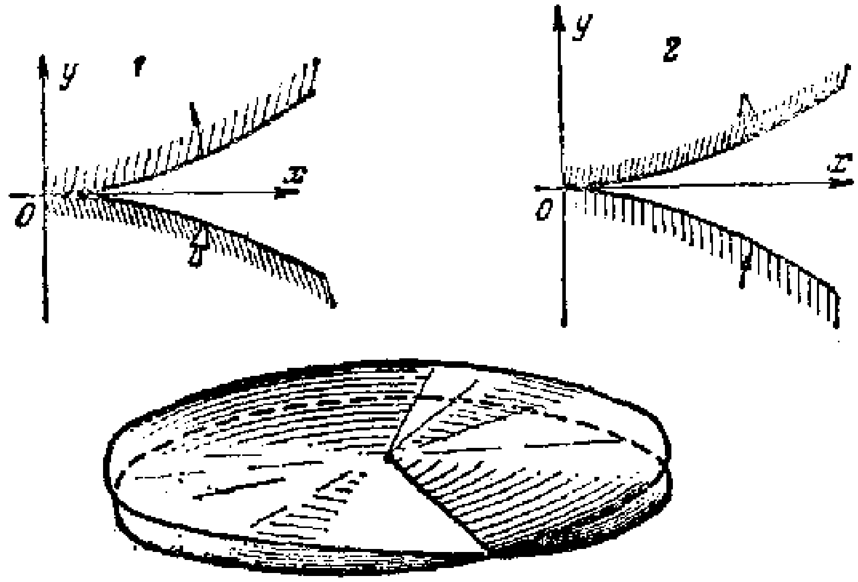


图 24

我们现在准备了两张同样的 z 平面, 把一张放在另一张上面, 并且沿 Ox 轴的正的部分切开. 假若 z 从 z_0 出发描过包围坐标轴原点的闭曲线 C, 则 $\sqrt{z}$ 即从一支变到另一支, 因而黎曼面上的点即从一叶变到另一叶. 假若我们将第一叶割线的下边与第二叶

割线的上边粘合起来，我们就可做到这一点．假若 z 第二次描过闭曲线 C，则 $\sqrt{z}$ 的值必然又回到出发时的值，因而黎曼面上的点必须又回到第一叶上出发时的位置．要想做到这一点，我们现在必须把第二叶的下边与第一叶的上边粘在一起．结果我们就得到了一张双叶面，它本身沿 Ox 轴的正的部分自相穿过．由图 24，我们可以得到这个面的一个表象，在该图中，我们画出的是 $z=0$ 这一点在这面上的一个邻域．

借以反应出多值函数的特征的多叶面可以用同样方法对于一切的多值函数构造出来．这个面的各叶在函数的支点周围相互结合在一起．看来，解析函数的性质与黎曼面的几何性质密切相关．引入这种面，不只是用来作为阐明多值函数的特征的辅助工具，而且对于研究解析函数的性质，以及发展研究解析函数的方法，皆具有根本性的作用．黎曼面就好象在复变数域与几何之间做成了一个桥梁，它不仅使得我们可以把非常深奥的函数的解析性质与几

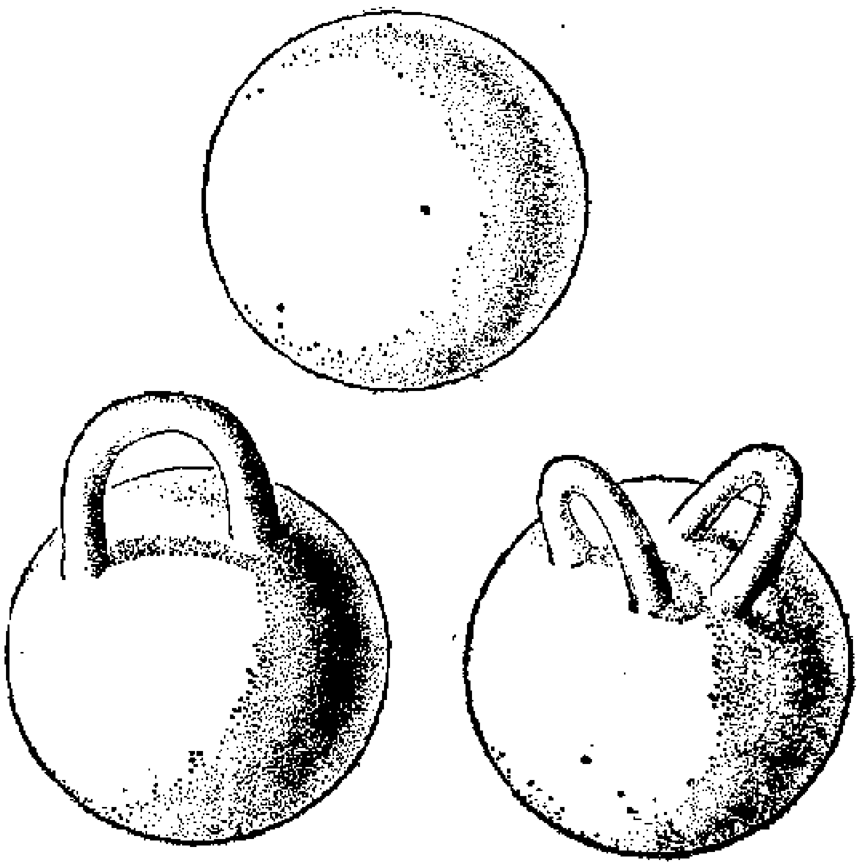


图 25

何联系起来，而且还刺激起几何学中一个新的分支——拓扑学——发展起来．拓扑学乃是研究图形在连续变形之下不变的几何性质的学科．

经过解方程

$$f(z, w)=0$$

得出的函数叫做代数函数，上式左边是 z 和 w 的多项式．代数函数论中有一个显著的例子，可以用来说明黎曼面的几何性质的重要性．代数函数的黎曼面在经过连续变换之后总可以变成一个球或一个带柄的球（图 25）．柄的数目就是这种面的特征性质．这个数目叫做该面的亏格，或代数函数的亏格，从亏格出发，我们就可得出该面．看来，代替函数的亏格决定了代数函数最重要的性质．

§6. 结　　论

由于解代数方程的缘故，产生了解析函数论．但在这种理论的发展过程中，随时都接触到一些新的问题．它使得我们有可能去阐明那些推动分析、力学、数学物理等向前发展的基本函数的实质．分析中许许多多的重要事实只有放在复域中来讨论才能彻底弄清楚．复变函数，解释成为流体力学和电动力学中最重要的一种向量场的特征，已经有了直接的物理说明，而且对于解决某些促进这两个科学部门发展的问题，也提供了重要的工具．函数论与传热性理论和弹性理论等等之间也发现有着联系．

微分方程论中的普遍问题以及解决这些问题的特殊方法，无论是过去或现在，皆广泛地依赖于复变函数论．在积分方程论和一般线性运算子理论中，自然也引进解析函数．解析函数论与几何学之间也出现密切的联系．函数论与数学中新的领域以及自然科学之间的这些日益增大的联系证实了这种理论的生命力，而且不断地丰富了它的题材．

在这篇短文中，我们不可能抱定一个目标，要把函数论中所有

各种各样的分支和联系都提供出一个概念．我们只企图通过一些属于函数论中某些基本方向的基础理论中的初等事实，说明这种理论中的若干问题和方向．许多重要的方向——与微分方程论、特殊函数论、椭圆函数论、自守函数论等等的关系，与三角级数的关系——和许多别的方向都完全没有接触到．在其他的情形，我们也不得不限于非常简短的说明．但我们盼望这篇短文能就复变函数论的性质和意义给读者们一个一般的印象．

文　献

通 俗 书 籍

B. Л. 岗恰罗夫，初等复变函数，初等数学百科全书，第三卷（有中文译本）．

A. И. 马库希维奇，复数与保角映象，青年数学丛书（有中文译本）．

教　科　书

M. A. 拉夫伦捷夫和 Б. B. 沙巴特，复变函数论方法，高等教育出版社，1957 年版．

书中叙述了这种理论的一些基本方法，同时还列举了大量的例题，说明这种理论在物理学中各种问题上的应用．

A. И. Маркушевич, Теория аналитических функций, Гостехиздат, 1950.

复变函数论的一本非常全面而有系统的教本．

И. И. Привалов, Введение в теорию функций комплексного переменного, Изд, 9-е, Гостехиздат, 1954.

E. Титчмарш, Теория функций, перев. с англ. Гостехиздат, 1951.

这不是一本有系统的课程，而只是介绍复变函数论中的若干部门．本书的基本方向是解析的，因而几何问题在这里很少注意．

综 合 性 文 章

A. Ф. 别尔曼特和 A. И. 马库希维奇，复变函数论，苏联数学 30 年（1917—1947）（有中文译本）．

赵民义　译

许孔时　校

第十章 素 数

§1. 数论研究什么和如何研究数论

整数 正如读者从开头第一章(第一卷)已经知道的那样，人类从最古老的年代就不得不和整数打交道，但是达到对无穷自然数列

$$1,\ 2,\ 3,\ 4,\ 5,\ \cdots \tag{1}$$

的了解，却经历了漫长的许多世纪. 现在在实际生活的各式各样的问题中，人们必须时常用到整数. 整数反映着自然界中大量的关系；在所有的与不连续量有关的问题中，整数是必需的数学工具。

同时整数在研究连续量时也起着重要的作用. 例如在数学分析中就研究解析函数展开成 x 整数次方的幂级数

$$f(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n+\cdots$$

的展开式。

所有的计算实质上是用整数进行的，对自动计算机和算术计算器的表面观察就立即可以看出此点，同样地，各种数学表，例如对数表也是如此. 对整数进行了运算之后，在固定的地方打上逗点，这就相当于造成十进位小数；这些小数与任何有理分数一样，表示两个整数的比. 其实，在计算任意的实数(例如 π)时，我们实际上是用有理分数来代替它的$\left(\text{例如，算做 } \pi=\frac{22}{7} \text{ 或 } \pi=3.14\right)$.

当算术研究建立数之运算法则的时候，数论研究自然数列(1)补充上零和负整数的更深刻的性质，数论是研究整数系的科学，在广泛的意义上说来，是研究利用整数按一定形式构成的数系的科学(详见本章 § 5). 当然，数论并不将整数孤立起来研究，而是在他们的相互联系中研究数的性质，研究如何用它们之间的关系来

确定它们.

数论的基本问题之一是研究一个数能否被另一个数整除的问题: 如果整数 a 被整数 b(不等于零)除的结果是整数, 就是如果

$$a=b\cdot c$$

(a, b, c 均为整数), 就说 a 被 b 除得尽或 b 除得尽 a. 如果整数 a 被整数 b 除的结果得到分数, 就说 a 不能被 b 除尽. 在实际生活中常常碰到数的可除性问题. 在数学分析的某些问题中, 数的可除性起着重要的作用. 例如, 若函数依 x 的整幂展开为

$$f(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n+\cdots \tag{2}$$

的展开式中全部奇系数(即指标不能被 2 除尽的项的系数)都等于零, 也即如果

$$f(x)=a_0+a_2x^2+\cdots+a_{2k}x^{2k}+\cdots,$$

则函数适合条件

$$f(-x)=f(x);$$

这样的函数称为偶函数, 它的图形关于纵坐标轴是对称的. 又若在展开式 (2) 中全部偶系数(即指标能被 2 除尽的项的系数)都等于零, 换句话说, 如果

$$f(x)=a_1x+a_3x^3+\cdots+a_{2k+1}x^{2k+1}+\cdots,$$

则

$$f(-x)=-f(x);$$

在这种情况下, 函数称为奇函数, 它的图形关于坐标原点是对称的. 例如

$$\sin x=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots \quad (\text{奇函数});$$

$$\cos x=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\cdots \quad (\text{偶函数}).$$

利用圆规和直尺能否做正 n 边形的几何问题决定于数 n 的算术性质[1].

大于 1 的整数之只有两个正整因数(即 1 及该数自己)者称为素数. 1 不算做素数, 因为它没有两个不同的正因数.

1) 参阅第一卷第四章, 289 页.

因此，素数就是这些数：

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, \cdots. \tag{3}$$

说明素数在数论中的基本作用的是数论的基本定理：每个整数 $n>1$ 可以表示成素数之乘积(因数可能有重复)，即可以表成

$$n=p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_k^{\alpha_k} \tag{4}$$

的形状，这里的 $p_1<p_2<\cdots<p_k$ 是一些素数，$\alpha_1, \alpha_2, \cdots, \alpha_k$ 是不小于1的整数，并且形如(4)式的 n 的表示法是唯一的.

数的性质之与将它表示为被加数之和的形状有关者称为**堆垒性质**；与将它表示为因子乘积的形状有关者称为**积性**. 数之堆垒性质与积性之间的联系是非常复杂的，它引出数论中一系列主要的问题.

整数是最简单明显的数学概念，以最紧密的形式联系着客观实际；由于这个原因，在数论中有困难问题出现的时候，常常引起这样一种情况：即在研究数论中深刻问题的时候常出现新的概念和发展强有力的方法，这些概念和方法往往不仅在数论中有意义，而且在别的数学部门中也有意义. 例如，"自然数列为无穷"这一概念对数学的全部发展有着巨大的影响，它反映出物质世界在空间和时间上为无限. 自然数列中的项各有其次序这一事实有着巨大的意义. 研究在整数上的演算就导出代数运算的概念，这在数学的许多部门中起着根本的作用.

在算术对象上首先形成的"算法"概念在数学中有着巨大的意义，"算法"是解问题的程序，它是基于重复使用某种严格的规定来解问题的程序；尤其是在机器数学中，"算法"的作用是很基本的. "算法"解问题的特点可以由求两个自然数 a 和 b 的最大公因数的欧几里得算法的例子中清楚地看出来.

令 $a>b$. 用 b 去除 a，若 b 不是 a 的因数，则得到不完全商 q_1 和余数 r_2，

$$a=bq_1+r_2,\quad 0<r_2<b. \tag{5_1}$$

其次，若 $r_2\neq0$，用 r_2 除 b，

$$b=r_2q_2+r_3,\quad 0<r_3<r_2. \tag{5_2}$$

然后，用 r_3 除 r_2，这样做下去一直到没有余数（即余数等于零）为止；因为非负整数 $r_2, r_3, \cdots$ 逐步减小，必然会达到这一步．令

$$r_{n-2}=r_{n-1}q_{n-1}+r_n, \tag{5_{n-1}}$$

$$r_{n-1}=r_nq_n, \tag{5_n}$$

那么，r_n 恰好就是 a 和 b 的最大公因数．实际上，如果两个整数 l 和 m 有公因数 d，则对二整数 h 与 k 而言，数 $hl+km$ 同样可被 d 除尽．我们设 a 和 b 的最大公因数等于 δ．从等式(5_1)可见 δ 是 r_2 的因数；由(5_2)可知 δ 是 r_3 的因数…；从(5_{n-1})可知 δ 是 r_n 的因数．但 r_n 本身是 a 和 b 的公因数，因为从(5_n)可知 r_n 是 r_{n-1} 的因数；从(5_{n-1}) r_n 是 r_{n-2} 的因数，等等．所以 δ 与 r_n 相同，于是求 a 和 b 的最大公因数的问题就解决了．在这里，对任何的 a 和 b 我们有某种同一类型的方法，它使得我们可以巧妙地求出所要的结果．以上是说明“算法”特征的一个例子．

数论曾对许多数学分支的发展发生过影响，如数学分析、几何、古典代数与近世代数、级数求和论、概率论及其他．

数论的方法　按照方法来说，数论可分为四个部分：初等数论、解析数论、代数数论和几何数论．

初等数论不求助于其他数学部门而研究整数的性质．例如由欧拉恒等式

$$\begin{aligned}(x_1^2+x_2^2+x_3^2+x_4^2)(y_1^2+y_2^2+y_3^2+y_4^2)\\ =(x_1y_1+x_2y_2+x_3y_3+x_4y_4)^2\\ +(x_1y_2-x_2y_1+x_3y_4-x_4y_3)^2\\ +(x_1y_3-x_3y_1+x_4y_2-x_2y_4)^2\\ +(x_1y_4-x_4y_1+x_2y_3-x_3y_2)^2\end{aligned} \tag{6}$$

出发，可以相当快地证明，每一个整数 $N>0$ 可以分解为四个整数平方的和，也即表示成

$$N=x^2+y^2+z^2+u^2$$

的形状，这里 x, y, z, u 都是整数[1]。

解析数论是用数学分析的工具解决数论问题．解析数论的基础是欧拉建立的，车比雪夫、狄里赫雷、黎曼、拉马努张、哈代、李特尔乌德和其他学者将它发展了．最有力的解析数论的方法是维诺格拉朵夫院士创造的，关于这一点将在下面谈到．数论的这一部分紧密地联系着象复变函数论这样广泛实用的数学分支，也和级数论、概率论及其他数学部门有关．

代数数论的基本概念是代数数的概念，代数数就是方程式

$$a_0x^n+a_1x^{n-1}+a_2x^{n-2}+\cdots+a_{n-1}x+a_n=0$$

的根，这里 a_0, a_1, a_2, $\cdots$, a_n 是平常的整数[2]．

在数论的这一方面有巨大贡献的是拉格朗日、高斯、库莫尔、佐洛塔辽夫、戴德金、盖尔芳特及其他一些人．

几何数论研究的基本对象是"空间格网"，也就是全部"整点"的组，"整点"是在给定的直线坐标系统（垂直的或斜的）中坐标全是整数的点．空间格网对几何学和结晶学有着巨大的意义，同时它的研究紧密地联系着数论中的重要问题（特别是二次型的算术理论，即系数和变数都是整数的二次型的理论）．几何数论方面的创始工作属于闵科夫斯基和沃洛诺伊．

必须指出，解析数论的方法不论在代数数论中和在几何数论中都有着重要的应用．这里特别应当指出的是，计算给定的区域中整点数目的方法对某些物理学的部门是很重要的．解决这一问题的道路是沃洛诺伊提出的，而他的解法又被维诺格拉朵夫发展了．

数论中解析方法有深刻的力量的原因就在于，存在于不连续的整数之间的联系由于经过连续量而引导出的新的联系，内容变得更为充实了．

必须着重指出，本章中研究的仅仅是从数论问题中选出的某些问题．

---

1) 这里，我们得到从有整数解的观点来研究不定方程的例子．

2) 如果 $a_0=1$，则代数数叫做代数整数．不是代数数的数叫做超越数．

§2. 如何研究与素数有关的问题

素数的个数无穷 当在研究素数序列(3)

$$2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19,\ \cdots$$

时，自然就发生这样的问题：这个序列是无穷的吗？任何整数可以表示成形状(4)这一事实还不能解决问题，因为指数 $a_1, \cdots, a_k$ 可以取无数多个值．所提出的问题的正面的回答还是欧几里得给出的，对于不可能有任何一个有限整数 k 使得素数仅有 k 个，其证明如下：

设 $p_1, p_2, \cdots, p_k$ 都是素数，则数

$$m=p_1p_2\cdots p_k+1$$

与每个整数一样，它大于 1，或者是素数，或者有素因子．但是 m 不能被素数 $p_1, p_2, \cdots, p_k$ 中的任何一个除尽，否则差 $m-p_1p_2\cdots p_k$ 就也被这个数除尽，而由于这个差等于 1，这是不可能的．因此，或者 m 本身是素数，或者 m 被另外的与 $p_1, p_2, \cdots, p_k$ 不同的素数 p_{k+1} 除得尽．这样一来，素数的集合就不可能为有限．

爱拉托斯芬的筛法 下面用来寻找不大于给定的自然数 N 的全部素数的"筛"法，要归功于公元前三世纪希腊的数学家爱拉托斯芬．写出从 1 到 N 的全部整数

$$1,\ 2,\ 3,\ 4,\ \cdots,\ N,$$

然后先划掉所有是 2 的倍数的数，2 不划掉，再划掉所有 3 的倍数，3 不划掉；再划掉所有 5 的倍数（所有 4 的倍数已经划掉了），5 不划掉；等等；1 也划掉；余下的数都是素数．必须指出，这种划掉数的手续只要继续进行到不发现任何小于 $\sqrt{N}$ 的素数为止，因为每一个不大于 N 的复合数（即非素数）必有不大于 $\sqrt{N}$ 的素因数．

在全部正整数的序列中研究素数的序列就会导出这样的结论：素数构成的规律性非常复杂．有这样的素数，例如 8004119 和 8004121，它们之间的差等于 2（这叫做孪生素数）；也有这样的素

数，它们彼此相距很远，例如 86629 和 86677，但它们之间一个素数也没有．同时，许多表表明，平均说来数越大素数就越稀少．

欧拉恒等式及其关于素数无穷的证明 十八世纪伟大的数学家俄国科学院院士欧拉开始研究了下面的自变量 $s>1$ 的函数：

$$\zeta(s)=1+\frac{1}{2^s}+\frac{1}{3^s}+\cdots+\frac{1}{n^s}+\cdots, \tag{7}$$

这函数现在记做 $\zeta(s)$．

正如我们从第一卷第二章知道的那样，实际上，给定的级数当 $s>1$ 时收敛（当 $s\leqslant 1$ 时发散）．欧拉指出了在素数论中起着非常重要作用的一个著名的恒等式：

$$\sum_{n=1}^{\infty}\frac{1}{n^s}=\prod_p\frac{1}{1-\frac{1}{p^s}}, \tag{8}$$

这里符号 $\prod_p$ 表示 $\frac{1}{1-\frac{1}{p^s}}$ 之积，p 跑过所有的素数．为了指出这个恒等式的证明过程，我们现指出当 $|q|<1$ 时有 $\frac{1}{1-q}=1+q+q^2+\cdots$，因此，

$$\frac{1}{1-\frac{1}{p^s}}=1+\frac{1}{p^s}+\frac{1}{p^{2s}}+\cdots.$$

将这些级数就不同的素数 p 连乘起来，并想到每个 n 唯一地分解为素数之积，我们就有

$$\prod_p\left(1+\frac{1}{p^s}+\frac{1}{p^{2s}}+\cdots\right)=1+\frac{1}{2^s}+\frac{1}{3^s}+\cdots+\frac{1}{n^s}+\cdots.$$

当然，对于严格的证明，必须论证我们所施行的极限手续是否可行，但这并没有什么困难．

从恒等式(8)可以得出这样的结果，即由全部素数的倒数做成的级数 $\sum_p\frac{1}{p}$ 是发散的（这对我们已知的素数不只有限个这一事实给出了一个新的证明），并且当 x 无限增大时，不大于 x 的素数的个数与 x 之比趋于零．

车比雪夫关于自然数列中素数分布的研究 正如现在所采用的那样，我们用 $\Pi(x)$ 表示不大于 x 的素数的个数：例如 $\Pi(10)=4$，因为不大于10的全部素数是 2, 3, 5, 7; $\Pi(\pi)=2$，因为不大于 π 的全部素数是 2 和 3. 我们已经提到过，

$$\lim_{x\to\infty}\frac{\Pi(x)}{x}=0.$$

比数 $\frac{\Pi(x)}{x}$ 减小的情况如何？或者换句话说，$\Pi(x)$按照什么规律增加呢？能不能找到一个比较简单而又是大家所熟悉的函数，它和 $\Pi(x)$ 的差别相对而言相当小. 著名的法国数学家莱让德尔据对于素数表的研究断言这样的函数是

$$\frac{x}{\ln x-A},\tag{9}$$

这里 $A=1.08\cdots$，但是没有给这一断言以证明. 高斯也研究了素数分布的问题，他做了这样的推测，即 $\Pi(x)$ 和 $\int_2^x\frac{dt}{\ln t}$ 的差别比较小.（我们注意，关系

$$\lim_{x\to\infty}\frac{\int_2^x\frac{dt}{\ln t}}{\frac{x}{\ln x}}=1,\tag{10}$$

成立，用分部积分法，再将得到的积分加以估计，就可验证这个关系式.）

在关于素数分布的最困难的问题上，从欧几里得的时候起第一个得到本质上推进的是车比雪夫. 在 1848 年，车比雪夫根据对 s 是实数时的欧拉函数 $\zeta(s)$ 的研究指出，若 $n>0$ 是任意大的数且 $\alpha>0$ 是任意小的数，则有任意大的数 x，对于它有

$$\Pi(x)>\int_2^x\frac{dt}{\ln t}-\frac{\alpha x}{\ln^n x},$$

也有任意大的数 x 使得

$$\Pi(x)<\int_2^x\frac{dt}{\ln t}+\frac{\alpha x}{\ln^n x},$$

这与高斯的推测完全符合．特别是当取 $n=1$，并计算 (10) 时，只要 $\Pi(x)/(x/\ln x)$ 的极限存在，车比雪夫建立了

$$\lim_{x\to\infty}\frac{\Pi(x)}{\frac{x}{\ln x}}=1. \tag{11}$$

车比雪夫也反驳了莱让德尔关于(9)式中所出现的给出 $\Pi(x)$ 的最佳逼近的 A 值所作的猜测，车比雪夫指出，这个值只可能是 $A=1$.

著名的法国数学家贝尔特兰在研究群论时需要证明下述的猜测，对于这一猜测，他就从 1 直到很大的 n，按照表和实际检验总说明它成立：若 $n>3$，则在 n 与 $2n-2$ 之间至少有一个素数．一直到 1850 年，贝尔特兰和其他数学家们证明这一猜测的全部企图都没得到什么结果，在 1850 年车比雪夫发表了对素数有贡献的第二个工作，其中不仅证明了前面指出过的猜测(贝尔特兰假设)，并且指出了当 x 充分大时，不等式

$$A_1<\frac{\Pi(x)}{\frac{x}{\ln x}}<A_2 \tag{12}$$

成立，这里 $0.92<A_1<1,\ 1<A_2<1.1.$

在 § 3 中我们将给车比雪夫方法以简单的叙述，实际上这方法所得出的是比车比雪夫自己的结果粗糙得多的结果．

车比雪夫的工作在许多数学家那里，特别是在西尔维斯特尔和庞卡莱那里，有着许多反应．在四十多年的时间内，许多学者研究如何改善车比雪夫不等式 (12) (即增大不等式左方的常数和减小右方的常数)．但是并不能证明极限

$$\lim_{x\to\infty}\frac{\Pi(x)}{\frac{x}{\ln x}}$$

存在(关于这一点，正如我们已经谈到的那样，从车比雪夫的工作中已经知道，若此极限存在则它等于 1)．

在 1896 年阿达马才从复变函数论的考虑证明了，车比雪夫在

他的研究中所引进的由等式

$$\Theta(x)=\sum_{p\leqslant x}\ln p$$

定义的函数 $\Theta(x)$ 适合条件

$$\lim_{x\to\infty}\frac{\Theta(x)}{x}=1, \tag{13}$$

由此已可不需任何补充假设就相当容易地得到关系式 (11) (这就是所谓素数分布定理).

阿达马是在十九世纪德国著名数学家黎曼的研究工作的基础上得到结果 (13) 的，黎曼研究了当变数 $s=\sigma+it$ 是复数时的欧拉函数 $\zeta(s)$ (7) [这正是车比雪夫研究变数是实数的 $\zeta(s)$ 的同时][1].

黎曼指出，由级数 (7),

$$\zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}$$

在半平面 $\sigma>1$ 上所定义的函数 $\zeta(s)$ 有这样的性质，即

$$\zeta(s)-\frac{1}{s-1}$$

是整超越函数 [当 $s\leqslant 1$ 时，级数 (7) 不再收敛，但 $\zeta(s)$ 在半平面 $\sigma\leqslant 1$ 上的值可以用解析开拓来定义] (参看第九章). 黎曼提出了一个猜测 ("黎曼假设") 说，$\zeta(s)$ 在区域 $0\leqslant\sigma\leqslant 1$ 中所有的零点的实部分都等于 $\frac{1}{2}$，就是说都在直线 $\sigma=\frac{1}{2}$ 上；这个猜测是否正确，是一个至今尚未解决的问题.

在直线 $\sigma=1$ 上 $\zeta(s)$ 没有零点这个事实的成立是证明 (13) 的重要步骤.

对 $\zeta(s)$ 性质的研究引起了有重要实际应用的整函数和半纯函数的严整理论的发展.

维诺格拉朵夫和他的学生们在素数论方面的工作 由 (10) 等式 (13) 可以写成

1) 在 1949 年，塞尔贝克成功地找到了素数分布定理的初等证明 (不依赖于复变函数论).

$$\lim_{x\to\infty}\frac{\Pi(x)}{\int_2^x \frac{dt}{\ln t}}=1 \tag{14}$$

之形式，在得出这等式之后，跟着就发生了一个问题，即函数 $\int_2^x \frac{dt}{\ln t}$ 表示 $\Pi(x)$ 精确到什么样的程度．这方面最好的结果是邱达柯夫在应用维诺格拉朵夫创造的三角和方法（关于这个方法将在 § 4 中谈到）的基础上得到的，这个方法也使邱达柯夫大大地降低了可以断定至少有一个素数的那个界限．早先曾经证明过如果研究序列

$$1^{250},\ 2^{250},\ 3^{250},\ \cdots,\ n^{250},\ (n+1)^{250},\ \cdots, \tag{15}$$

则从某一个 $n=n_0$ 开始，在相邻两项之间，即 n^{250} 与 $(n+1)^{250}$ 之间，至少有一个素数．

我们注意，由二项式公式有

$$(n+1)^{250}-n^{250}>250n^{249},$$

所以这个差是很大的．邱达柯夫成功地用序列

$$1^4,\ 2^4,\ 3^4,\ \cdots,\ n^4,\ (n+1)^4,\ \cdots \tag{16}$$

代替了序列(15)，显然这比序列(15)精密多了，然而也是从某一个 $n=n_0$ 开始在相邻两项之间即 n^4 与 $(n+1)^4$ 之间至少含有一个素数．后来这个结果又改进了，用 3 代替了 4．

如果 k 和 l 互素（即没有公因数）且都大于 1，则在以 $kt+l$ 为一般项的等差级数中含有无穷多个素数．这个事实是欧几里得结果的推广，在十九世纪为狄里赫雷所证明．在级数 $kt+l$ 中最小素数所显然不超过的那个界数的情况如何呢？列宁格勒的数学家林尼克证明了有绝对常数 C 存在，C 有这样的性质，即在任何级数 $kt+l$（k 与 l 互素）中一定至少可以找到一个小于 k^C 的素数．尤其是从原则上看，林尼克已几乎完全解决了这一个提出多年的等差级数中最小素数的问题，下一步的研究只能是减小常数 C 的值．与 $\zeta(s)$ 和其他更一般的函数之零点有关的问题方面很重要的研究工作也是属于林尼克的．

正如我们已经提到过的那样，关于素数分布问题方面，最好的结果是借助维诺格拉朵夫三角和估值的方法才得到的.

形如
$$\sum_{A<x<B} e^{2\pi i f(x)}$$
的和数称为**三角和**，这里 $f(x)$ 是 x 的某一个实函数，并且 x 跑过在 A 与 B 之间的全部整数，或者跑过这些整数的某一部分，例如 A 与 B 之间的素数. 因为当 Z 是实数时 $e^{2\pi i z}$ 的绝对值等于 1，并且一些项之和的绝对值不大于这些项的绝对值之和，所以

$$\left|\sum_{x=1}^{p} e^{2\pi i f(x)}\right| \leqslant p. \tag{17}$$

这种显而易见的估值在许多情况下得到了本质上的改善；在这方面决定性的步骤是维诺格拉朵夫做的. 为了确定起见，令 $f(x)$ 是多项式

$$f(x)=\alpha_n x^n+\alpha_{n-1}x^{n-1}+\cdots+\alpha_1 x+\alpha_0.$$

如果所有的 α 都是整数，则当 x 是整数时 $e^{2\pi i f(x)}=1$，在这种情况下估值(17)当然不能改进. 但如果 $\alpha_1, \cdots, \alpha_n$ 不全是整数，则正如维诺格拉朵夫证明的那样，估值 (17) 可以借助于将这些系数中的某些个用有理分数接近的办法来弄得更精确，这些分数的分母不大于某个极限$\Big($可以证明，0 与 1 之间的任一 α 皆可以表为 $\alpha=\dfrac{a}{q}+z$ 的形状，这里 a 和 q 是互素的整数，$q\leqslant\tau$，$|z|\leqslant\dfrac{1}{q\tau}$，$\tau$ 是事先给定的大于 1 的整数$\Big)$.

维诺格拉朵夫用他创造的三角和方法解决了数论中一系列最困难的问题. 特别是在 1937 年，维诺格拉朵夫解决了著名的古德巴赫问题，他证明了每一个充分大的奇数 N 可以表示成三个素数之和的形状

$$N=p_1+p_2+p_3. \tag{18}$$

这个问题是 1742 年欧拉和俄国科学院的另一位院士古德巴赫的通信中提出来的，在将近两百年的时期中，纵然有许多卓越的数学家努力企图解决这个问题，但问题并未解决.

正如我们已经看到的那样，按照等式(4)，在整数的乘法表示中素数起着基本的作用，而等式(18)则将奇数用素数以加法表示。不难看出，从(18)可以导出充分大的偶数可以表示为不超过四个素数之和的形状[1]。这样一来，维诺格拉朵夫-古德巴赫定理就确定了数之堆垒性质与积性之间的最深刻的联系。

维诺格拉朵夫所创造的三角和方法的作用并不限于数论一方面。特别这个方法在函数论和概率论方面也起着重要的作用。关于维诺格拉朵夫方法的某些介绍可以在本章§4中得到。

希望更详细地了解这一方法的读者可以去读维诺格拉朵夫著的"数论中的三角和方法"一书，但是要先读维诺格拉朵夫著的"数论基础"一书。

§3. 关于车比雪夫方法

车比雪夫函数 Θ 及其估值 现在我们给计算给定范围内素数个数的车比雪夫方法以简单的说明。为了写起来简短起见，我们约定使用下面的记号：如果 B 是某个可以无限增大的正变量，A 是另一个变量使得 $|A|$ 的增大不快于 CB，此处 C 是一个正的常数$\left(\text{准确些说，如果存在常数 } C>0 \text{ 使得从某个时候起永远有 } \frac{|A|}{B}\leqslant C\right)$，则我们就写

$$A=O(B),$$

这个通常读做："A 的阶是 B"。例如 $\sin x=O(1)$，

因为永远有

$$\frac{|\sin x|}{1}\leqslant 1;$$

正是这样，$5x^3\cos 2x=O(x^3)$。

我们也用 $[x]$ 来记 x 的整数部分，即不大于 x 的最大整数；例如

1) 关于欧拉猜测是否成立的问题，即每个充分大的偶数 N 表为两个素数之和的问题，至今尚未解决。

$$[\pi]=3,\ [5]=5,\ [-1.5]=-2,\ [0.999]=0.$$

我们现在提出下面的问题：令 p 是素数，n 是自然数，而 $n!$ 象通常那样表示乘积 $1\cdot 2\cdot 3\cdots n$（顺便可以看出，当 n 增加时，量 $n!$ 增加得非常之快）。能除得尽 $n!$ 的素数 p 的最大方次 a 是多少呢？

在数 $1, 2, \cdots, n$ 当中被 p 除得尽的数有 $\left[\frac{n}{p}\right]$ 个，在这些数中也被 p^2 除得尽的数有 $\left[\frac{n}{p^2}\right]$ 个；在后面这些数中被 p^3 除得尽的数有 $\left[\frac{n}{p^3}\right]$ 个，等等．由此不难看出

$$a=\left[\frac{n}{p}\right]+\left[\frac{n}{p^2}\right]+\left[\frac{n}{p^3}\right]+\cdots$$

$\left(\text{此处级数本身会中断的，因为}\left[\frac{n}{p^s}\right]>0\text{ 只当 }n\geqslant p^s\text{ 时成立}\right)$．实际上，在上面的和数中，乘积 $1\cdot 2\cdot 3\cdots n$ 的每一个因数是这样的，即除得尽这一因数的 p 的方次高到 p^m 的话，这个因数就恰恰算了 m 次；因为它是 p 的倍数，p^2 的倍数，p^3 的倍数，…，最后它是 p^m 的倍数．

从得到的结果和任一自然数可以表为(4)式的形状就可导出，$n!$ 是形如

$$p^{\left[\frac{n}{p}\right]+\left[\frac{n}{p^2}\right]+\left[\frac{n}{p^3}\right]+\cdots}$$

的 p 的方幂的乘积，这里要取全部 $p\leqslant n$ 的素数．因此 $\ln(n!)$ 就将是这些素数方幂的对数和，可以简写为下面的形状：

$$\ln n!=\sum_{p\leqslant n}\left(\left[\frac{n}{p}\right]+\left[\frac{n}{p^2}\right]+\left[\frac{n}{p^3}\right]+\cdots\right)\ln p. \tag{19}$$

我们来化减等式(19)．因为函数 $y=\ln x$ 是递增函数，于是

$$\ln m=\ln m\int_m^{m+1}dx<\int_m^{m+1}\ln x\,dx<\ln(m+1)\int_m^{m+1}dx=\ln(m+1)$$

（在图 1 中尤其明显）．

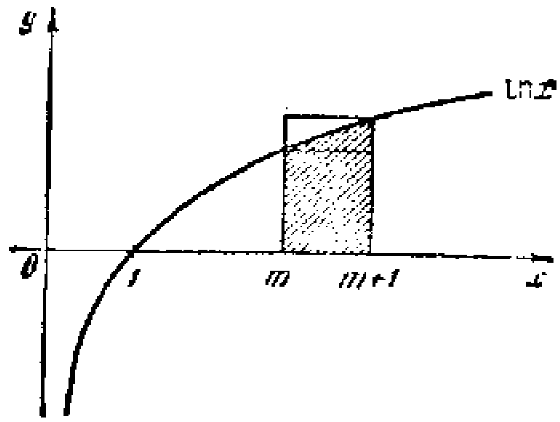


图 1

因此

$$\ln n! = \ln 1 + \ln 2 + \cdots + \ln n$$
$$< \int_1^2 \ln x\,dx + \int_2^3 \ln x\,dx + \cdots$$
$$+ \int_{n-1}^{n} \ln x\,dx + \ln n = \int_1^n \ln x\,dx + \ln n,$$

另一方面有

$$\ln n! > \ln 1 + \int_1^2 \ln x\,dx + \cdots + \int_{n-2}^{n-1} \ln x\,dx + \int_{n-1}^{n} \ln x\,dx$$
$$= \int_1^n \ln x\,dx.$$

利分用部积分公式，我们有

$$\int_1^n \ln x\,dx = [x\cdot\ln x]_1^n - \int_1^n x\cdot\frac{1}{x}\,dx = n\ln n - (n-1).$$

于是

$$n\ln n - n + 1 < \ln n! < n\ln n - n + 1 + \ln n,$$

由此得出

$$\ln n! = n\ln n + O(n). \tag{20}$$

我们指出 $\ln n = O(n)$； 并且当 $n\to\infty$ 时，$\ln n$ 的增大要比 n 的任意正方幂的增大为慢，就是说对任一常数 $a>0$，有

$$\lim_{n\to\infty} \frac{\ln n}{n^a} = 0 \tag{21}$$

因为按照未定式展开的法则[参看第二章(第一卷)140 页]有

$$\lim_{n\to\infty}\frac{\ln n}{n^\alpha}=\lim_{n\to\infty}\frac{\frac{1}{n}}{\alpha n^{\alpha-1}}=\frac{1}{\alpha}\lim_{n\to\infty}\frac{1}{n^\alpha}=0.$$

其次，我们有

$$\begin{aligned}\sum_{p\leqslant n}\left(\left[\frac{n}{p^2}\right]+\left[\frac{n}{p^3}\right]+\cdots\right)\ln p&\leqslant\sum_{p\leqslant n}\left(\frac{n}{p^2}+\frac{n}{p^3}+\cdots\right)\ln p\\&=\sum_{p\leqslant n}\frac{n\ln p}{p^2\left(1-\frac{1}{p}\right)}<2n\sum_{p\leqslant n}\frac{\ln p}{p^2}<2n\sum_{m=1}^{\infty}\frac{\ln m}{m^2}\\&=2nC_0=O(n),\end{aligned}\tag{22}$$

这里 C_0 是收敛级数 $\sum_{m=1}^{\infty}\frac{\ln m}{m^2}$ 的和．例如当 $\alpha=\frac{1}{2}$ 时由 (21) 的计算和用比较法则及所谓级数收敛的积分法则［参看第二章（第一卷）§ 14］，就确定了这个级数的绝对收敛性．由于(20)和(22)，等式(19)可以导出

$$\sum_{p\leqslant n}\left[\frac{n}{p}\right]\ln p=n\ln n+O(n).\tag{23}$$

现在我们来研究车比雪夫引入的函数

$$\Theta(n)=\sum_{p\leqslant n}\ln p\tag{24}$$

（不大于 n 的全体素数之乘积的对数）．

我们可以把等式(23)写成这样：

$$\Theta\left(\frac{n}{1}\right)+\Theta\left(\frac{n}{2}\right)+\Theta\left(\frac{n}{3}\right)+\Theta\left(\frac{n}{4}\right)+\cdots=n\ln n+O(n).\tag{25}$$

实际上，每个给定的 $\ln p$ 在所有那些形状是 $\Theta\left(\frac{n}{s}\right)$ 的和中都有，这里 $p\leqslant\frac{n}{s}$，也就是 $s\leqslant\frac{n}{p}$．这些和 $\Theta\left(\frac{n}{s}\right)$ 的个数是 $\left[\frac{n}{p}\right]$．

当 n 不是整数时，等式(25)也是正确的．为了使人信服起见，显然指出这一点就足够了，即对所有适合条件 $n<x<n+1$ 的 x 它正确：而为此指出这一点就足够了，即当用 x 代替 n 时(25)式左方不变，而右方第一项只可以增加一个 $O(n)$．但第一点可从这种情况得出，即在如此替换之后左方各项中一项也不增加(仅当 n 之增加不小于 1 时才能有相反的情形)，当然也不减少．第二点

从那种情况可以得到，即按照函数增量的公式[参看第二章（第一卷)]

$$f(x)-f(a)=(x-a)f'(\xi),\quad a<\xi<x,$$

我们有 $x\ln x-n\ln n=(x-n)(\ln\xi+1),\quad n<\xi<x,$

并且因为 $0<x-n<1$，所以得到的等式的右方小于 $\ln(n+1)+1=O(n)$. 在(25)中用 $\frac{n}{2}$ 代替 n 得到一个等式，将这等式的每一项用 2 乘，从(25)中减去它就有

$$\Theta\left(\frac{n}{1}\right)+\Theta\left(\frac{n}{2}\right)+\Theta\left(\frac{n}{3}\right)+\Theta\left(\frac{n}{4}\right)+\cdots=n\ln n+O(n),$$

$$2\Theta\left(\frac{n}{2}\right)+2\Theta\left(\frac{n}{4}\right)+\cdots=2\cdot\frac{n}{2}\cdot\ln\frac{n}{2}+O(n),$$

$$\Theta\left(\frac{n}{1}\right)-\Theta\left(\frac{n}{2}\right)+\Theta\left(\frac{n}{3}\right)-\Theta\left(\frac{n}{4}\right)+\cdots=n\ln 2+O(n)<C_1n,$$

这里 C_1 是某一个正的常数. 因为差 $\Theta\left(\frac{n}{3}\right)-\Theta\left(\frac{n}{4}\right)$, $\Theta\left(\frac{n}{5}\right)-\Theta\left(\frac{n}{6}\right)$, $\cdots$ 不能是负的，所以 $\Theta\left(\frac{n}{1}\right)-\Theta\left(\frac{n}{2}\right)$不大于左方的全体. 因此从上面得到的不等式就有

$$\Theta\left(\frac{n}{1}\right)-\Theta\left(\frac{n}{2}\right)<C_{1n}.$$

在这里有 n 的地方放上数 $\frac{n}{2}$, $\frac{n}{4}$, $\cdots$. 我们也得到

$$\Theta\left(\frac{n}{2}\right)-\Theta\left(\frac{n}{4}\right)<C_1\cdot\frac{n}{2},$$

$$\Theta\left(\frac{n}{4}\right)-\Theta\left(\frac{n}{8}\right)<C_1\cdot\frac{n}{4},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots,$$

由此，考虑到在 k 充分大时 $\left(\text{当}\ \frac{n}{2^k}<2\ \text{时}\right)\Theta\left(\frac{n}{2^k}\right)=0$，逐项相加，我们得到

$$\Theta(n)<C_1\left(n+\frac{n}{2}+\frac{n}{4}+\cdots\right)=2C_1n. \tag{26}$$

其次，研究等式(23)，我们发现

$$0\leqslant\sum_{p\leqslant n}\frac{n}{p}\ln p-\sum_{p\leqslant n}\left[\frac{n}{p}\right]\ln p\leqslant\sum_{p\leqslant n}\ln p$$
$$=\Theta(n)\leqslant 2C_1 n=O(n),$$

由等式(23)就给出

$$\sum_{p\leqslant n}\frac{n}{p}\ln p=n\ln n+O(n),$$

$$\sum_{p\leqslant n}\frac{\ln p}{p}=\ln n+\theta C,\tag{27}$$

这里 C 是大于零的常数，θ 是与 n 有关的一个数，$|\theta|\leqslant 1$.

在固定区间中素数的总数之估计 现在我们指出，可以选择某个正的常数 M，使得如果 n 充分大，则在 n 与 Mn 之间有任意多的素数 p. 就是说，对在区间 $n<p\leqslant Mn$ 内素数的个数 T，我们可以建立一些简单的不等式. 显然，

$$\sum_{n<p\leqslant Mn}\frac{\ln p}{p}=\sum_{p\leqslant Mn}\frac{\ln p}{p}-\sum_{p\leqslant n}\frac{\ln p}{p}.\tag{28}$$

从等式(27)，用 Mn 代替 n，我们得到

$$\sum_{p\leqslant Mn}\frac{\ln p}{p}=\ln(Mn)+\theta' C=\ln M+\ln n+\theta' C,\tag{29}$$

这里 $|\theta'|\leqslant 1$；因而由等式(28)，(29)和(27)有

$$\sum_{n<p\leqslant Mn}\frac{\ln p}{p}=\ln M+\theta' C-\theta C=\ln M+2\theta_0 C,$$

这里 $|\theta_0|\leqslant 1$，也就是

$$\ln M-2C\leqslant\sum_{n<p\leqslant Mn}\frac{\ln p}{p}\leqslant\ln M+2C.\tag{30}$$

但是另一方面，因为当 $x>e$ 时函数 $y=\frac{\ln x}{x}$ 是递减的$\left(\text{因为当}\ln x>1\text{ 即 }x>e\text{ 时，}y'=\frac{1-\ln x}{x^2}>0\right)$，所以当 $n\geqslant 3$ 时有

$$T\frac{\ln Mn}{Mn}\leqslant\sum_{n<p\leqslant Mn}\frac{\ln p}{p}\leqslant T\frac{\ln n}{n},$$

从这里由(30)就得到

$$T\frac{\ln n}{n}>\ln M-2C\tag{31}$$

和 $$T\frac{\ln(Mn)}{Mn}<\ln M+2C. \tag{32}$$

现在我们取常数 M 使得(31)式右方等于1,

$$\ln M-2C=1,$$

也就是 $M=e^{2C+1}$, 并且我们令

$$L=M(\ln M+2C).$$

那么, 对于在 n 和 Mn 之间的素数的个数 T, 从(31)和(32)我们就得到不等式

$$\frac{n}{\ln n}<T<L\frac{n}{\ln n}, \tag{33}$$

这些不等式就是我们所想要建立的. 因为当 n 无限增大时 $\frac{n}{\ln n}\to\infty$, 所以在此情况下 $T\to\infty$.

§4. 维诺格拉朵夫方法

维诺格拉朵夫方法在解决古德巴赫问题时的应用 在本节中, 我们打算在个别的例子上给维诺格拉朵夫方法某种介绍, 例如用它来解决关于将奇数表示成三个素数之和的形状的古德巴赫问题.

将 N 表为三个素数之和的表示法的个数用积分的形状来表示 令 N 是一个充分大的奇数. 用 $I(N)$ 记 N 表为三个素数之和的表示法的数目, 换句话说, $I(N)$ 就是方程式

$$N=p_1+p_2+p_3 \tag{34}$$

当 p_1, p_2, p_3 都是素数时的解的数目.

如果确定了 $I(N)>0$, 则古德巴赫问题就解决了. 维诺格拉朵夫方法不仅使这个事实(当 N 充分大时)能够成立, 并且使得能够找到 $I(N)$ 的渐近表示法.

$I(N)$ 可以写成下面的形状:

$$I(N)=\sum_{p_1<N}\sum_{p_2<N}\sum_{p_3<N}\int_0^1 e^{2\pi i(p_1+p_2+p_3-N)\alpha}d\alpha, \tag{35}$$

这里对不大于 N 的素数进行求和. 事实上，当整数 $n\neq 0$ 时，有

$$\int_0^1 e^{2\pi in\alpha}d\alpha=\frac{1}{2\pi ni}\Big[e^{2\pi in\alpha}\Big]_0^1=\frac{1}{2\pi ni}(e^{2\pi ni}-e^0)=0,$$

因为

$$e^{2\pi in}=\cos 2\pi n+i\sin 2\pi n=1;$$

如果 $n=0$, 则

$$\int_0^1 e^{2\pi in\alpha}\,d\alpha=\int_0^1 d\alpha=1.$$

这样一来，每次当素数 p_1, p_2, p_3 之和给出 N 时，在(35)中的记号之下，积分就变成1；而当和数 $p_1+p_2+p_3\neq N$ 时，这个积分等于零；这就指出了等式(35)的正确性.

因为 $e^{2\pi ia}\cdot e^{2\pi ib}=e^{2\pi i(a+b)}$，并且一些项之和的积分等于这些项的积分之和，于是从等式(35)就得出

$$I(N)=\int_0^1\Big(\sum_{p\leqslant N}e^{2\pi i\alpha p}\Big)^3 e^{-2\pi i\alpha N}\,d\alpha.$$

引入记号

$$T_\alpha=\sum_{p\leqslant N}e^{2\pi i\alpha p};\qquad(36)$$

于是

$$I(N)=\int_0^1 T_\alpha^3 e^{-2\pi i\alpha N}d\alpha.\qquad(37)$$

将积分区间分为主要区间和辅助区间 令 h 是某个适当选定的与 N 有关的数，它与 N 同时无限增大，但它小于 N 甚至也小于 $\sqrt[3]{\frac{N}{2}}$, 令 $\tau=\frac{N}{h}$. 因为在(37)中的被积函数有周期，周期等于 1, 所以在 (37) 中积分区间可以用从 $-\frac{1}{\tau}$ 到 $1-\frac{1}{\tau}$ 的区间代替，因此

$$I(N)=\int_{-\frac{1}{\tau}}^{1-\frac{1}{\tau}}T_\alpha^3 e^{-2\pi i\alpha N}d\alpha.\qquad(38)$$

我们现在研究所有正的不可约的并且分母不大于 h 的分数 $\frac{a}{q}$, 并从区间 $-\frac{1}{\tau}\leqslant\alpha\leqslant 1-\frac{1}{\tau}$ 中分出来对应于这些分数的“主要”区间

$$\frac{a}{q}-\frac{1}{\tau}\leqslant\alpha\leqslant\frac{a}{q}+\frac{1}{\tau};\qquad(39)$$

当 N 充分大时，可以证明[1]这些区间没有公共点．这样一来，区间 $-\frac{1}{\tau}\leqslant\alpha\leqslant1-\frac{1}{\tau}$ 就分成许多主要区间和"辅助"区间．

将 $I(N)$ 表示成为两项之和的形状

$$I(N)=I_1(N)+I_2(N),\tag{40}$$

这里 $I_1(N)$ 表示在主要区间上的积分的和，而 $I_2(N)$ 表示在辅助区间上的积分的和．正如在今后将要看到的那样，当奇数 N 无限增大时，$I_1(N)$ 也无限增大，并且

$$\lim_{N\to\infty}\frac{I_2(N)}{I_1(N)}=0.\tag{41}$$

这样一来，由于 (40)，奇数 N 表为三个素数之和的表示法的数目就与 N 同时无限增大，特别是这就对所有充分大的奇数 N 证明了古德巴赫猜测．

在主要区间上的积分的表示法　令 α 属于主要区间中的一个区间，按照(39)，$\alpha=\frac{a}{q}+z$，并且 $1\leqslant q\leqslant h$，$|z|\leqslant\frac{1}{\tau}$．我们把对不大于 N 的全部素数普遍求和的和数(36)

$$T_\alpha=\sum_{p\leqslant N}e^{2\alpha i p\alpha}=\sum_{p\leqslant N}e^{2\pi i\left(\frac{a}{q}+z\right)p}$$

划分成许多形如

$$T_{\alpha,M}=\sum_{M<p<M'}e^{2\pi i\left(\frac{a}{q}+z\right)p}$$

---

1) 如果在点 $\frac{a_1}{q_1}$ 和 $\frac{a_2}{q_2}$ 的近旁两个这样的区间相交，则在公共点处等式

$$\frac{a_1}{q_1}+\frac{\theta_1}{\tau}=\frac{a_2}{q_2}+\frac{\theta_2}{\tau}$$

成立，这里 $|\theta_1|\leqslant1$，$|\theta_2|\leqslant1$，或者换个样子，

$$\frac{a_1q_2-a_2q_1}{q_1q_2}=\frac{\theta_2-\theta_1}{\tau}.$$

但最后等式左方的绝对值不小于 $\frac{1}{q_1q_2}$，即大于 $\frac{1}{h^2}$，而右方不大于 $\frac{2}{\tau}$，即小于 $\frac{2h}{N}$．如果最后的等式成立，则由它可以得出不等式

$$\frac{1}{h^2}<\frac{2h}{N},$$

这与 h 的选择相矛盾．

的部分和$T_{\alpha,M}$，这里M'是由这样的计算选定的，就是使得$e^{2\pi i\alpha p}$与$e^{2\pi i\alpha M}$差别"小"(考虑到只是要了解维诺格拉朵夫方法而不给出古德巴赫——维诺格拉朵夫定理的证明，在这里及今后我们不确定"差别小"所表示的意思；事实上在维诺格拉朵夫的证明中谈到的是与大量计算有关的严格地确定的许多不等式)．这样一来，

$$T_{\alpha,M}\approx e^{2\pi iM\varepsilon}\sum_{M<p<M'}e^{2\pi i\frac{a}{q}p}=e^{2\pi iM\varepsilon}T_{\frac{a}{q},M},\tag{42}$$

这里符号"≈"指的是最后关系式的左方与它的平均部分差别"小"．其次，从和数

$$T_{\frac{a}{q},M}=\sum_{M<p<M'}e^{2\pi i\frac{a}{q}p}\tag{43}$$

中分出每个和数$T_{\frac{a}{q},M'l}$，它对适合关系式$M\leqslant p_e<M'$并属于算术级数$qx+l$的素数p_e普遍求和，这里l取从0到$q-1$中与q互素的所有的值．但是

$$e^{2\pi i\frac{a}{q}pl}=e^{2\pi ka+2\pi i\frac{a}{q}l}=e^{2\pi i\frac{a}{q}l},$$

因此

$$T_{\frac{a}{q},M'l}=e^{2\pi i\frac{a}{q}l}\cdot\Pi(M,\ M',\ l),\tag{44}$$

这里$\Pi(M,\ M',\ l)$是适合条件$M\leqslant p<M'$并属于算术级数$qx+l$的素数的个数．在推广不大于x的素数个数$\Pi(x)$的公式(14)时，对于与差$M'-M$比较起来"小"的数值q，确定了$\Pi(M,\ M',\ l)$与$\frac{1}{\varphi(q)}\int_M^{M'}\frac{dx}{\ln x}$差别不大，这里$\varphi(q)$是欧拉函数．这是一个算术函数(即对自然数$q$定义的函数)，它表示不大于$q$且与$q$互素的正整数的个数．这样一来，由(44)可以得到

$$T_{\frac{a}{q},M'l}\approx e^{2\pi i\frac{a}{q}l}\cdot\frac{1}{\varphi(q)}\int_M^{M'}\frac{dx}{\ln x}.\tag{45}$$

在表示式中，在(45)的右方只是第一个因子与l有关，就是与算术级数$qx+l$有关(现在我们认为q是固定的)．在对l求和之后，得到

$$T_{\frac{a}{q},M} \approx \frac{1}{\varphi(q)} \int_M^{M'} \frac{dx}{\ln x} \sum_l e^{2\pi i \frac{a}{q} l},$$

其次，由(42)有

$$T_{\alpha,M} \approx e^{2\pi i M z} \cdot \frac{1}{\varphi(q)} \int_M^{M'} \frac{dx}{\ln x} \cdot \sum_l e^{2\pi i \frac{a}{q} l}, \tag{46}$$

并且因为

$$e^{2\pi i M z} \int_M^{M'} \frac{dx}{\ln x} \approx \int_M^{M'} \frac{e^{2\pi i z x}}{\ln x} dx,$$

于是用关系式

$$T_{\alpha,M} \approx \int_M^{M'} \frac{e^{2\pi i z x}}{\ln x} dx \cdot \frac{1}{\varphi(q)} \sum_l e^{2\pi i \frac{a}{q} l} \tag{47}$$

代替(46)．对 M 求和之后就确定

$$T_\alpha \approx \int_2^N \frac{e^{2\pi i z x}}{\ln x} dx \cdot \frac{1}{\varphi(q)} \sum_l e^{2\pi i \frac{a}{q} l}. \tag{48}$$

在关系式(48)右方的和数

$$\sum_l e^{2\pi i \frac{a}{q} l}$$

可用算术函数 $\mu(q)$ 表式出来，这里对不大于 q 且与 q 互素的自然数 l 求和，$\mu(q)$ 是用下述方式定义的：如 q 被大于 1 的整数的平方除得尽，则 $\mu(q)=0$；$\mu(1)=1$；如 $q=p_1p_2\cdots p_n$ 且 $p_1, p_2, \cdots, p_n$ 是不同的素数，则 $\mu(q)=(-1)^n$．当 a 与 q 互素时，就是

$$\sum_l e^{2\pi i \frac{a}{q} l} = \mu(a). \tag{49}$$

所以关系式(48)可以写成

$$T_\alpha \approx \frac{\mu(q)}{\varphi(q)} \int_2^N \frac{e^{2\pi i z x}}{\ln x} dx.$$

由于 $\mu^3(q)=\mu(q)$，故

$$T_\alpha^3 \approx \frac{\mu(q)}{(\varphi(q))^3} \left(\int_2^N \frac{e^{2\pi i z x}}{\ln x} dx \right)^3. \tag{50}$$

根据 $I_1(N)$ 的定义，我们有

$$I_1(N) = \sum_{1 < q < h} \sum_a \int_{\frac{a}{q}-\frac{1}{\tau}}^{\frac{a}{q}+\frac{1}{\tau}} T_\alpha^3 e^{-2\pi i \alpha N} d\alpha, \tag{51}$$

这里当 q 给定时对所有小于 q 的非负的 a 进行求和. 因为 $\alpha=\frac{a}{q}+z$, 于是由(50)有

$$I_1(N)\approx\sum_{1\le q<h}\frac{\mu(q)}{(\varphi(q))^3}\sum_a e^{-2\pi i\frac{a}{q}N}\int_{-\frac{1}{\tau}}^{\frac{1}{\tau}}\left(\int_2^N\frac{e^{2\pi izx}}{\ln x}dx\right)^3e^{-2\pi izN}dz. \tag{52}$$

引入记号

$$R(N)=\int_{-\frac{1}{\tau}}^{\frac{1}{\tau}}\left(\int_2^N\frac{e^{2\pi izx}}{\ln x}dx\right)^3e^{-2\pi izN}dz. \tag{53}$$

由关系式(52)得到

$$I_1(N)\approx R(N)\sum_{1\le q<h}\frac{\mu(q)}{(\varphi(q))^3}\sum_a e^{-2\pi i\frac{a}{q}N}. \tag{54}$$

必须注意那个情况, 即 $R(N)$ 是某个能进行概算的解析表示; 就是有

$$R(N)\approx\frac{N^2}{2(\ln N)^3}. \tag{55}$$

在关系式(54)右方因数 $R(N)$ 旁边的表示式与无穷极数和

$$S(N)=\sum_{q=1}^{\infty}\frac{\mu(q)}{(\varphi(q))^3}\sum_a e^{-2\pi i\frac{a}{q}N} \tag{56}$$

相差不大, 于是由(54)和(55)就能建立

$$I_1(N)\approx\frac{N^2}{2(\ln N)^3}S(N), \tag{57}$$

或者更准确些说,

$$I_1(N)=\frac{N^2}{2(\ln N)^3}(S(N)+\gamma_1(N)), \tag{58}$$

这里

$$\lim_{N\to\infty}\gamma_1(N)=0. \tag{59}$$

我们指出, 算术表示式 $S(N)$ 可以表示成

$$S(N)=C\prod_p\left(1-\frac{1}{p^2-3p+3}\right) \tag{60}$$

的样子, 这里 C 是某个常数, 是过 N 的全部素因子的某个乘积, 并

且计算表明

$$S(N)>0.6. \tag{61}$$

对在辅助区间上的积分的估计 现在我们转到对在辅助区间上许多积分之和 I_2 的估计. 因为积分的绝对值不大于被积函数的绝对值的积分，并且当 αN 是实数时 $|e^{-2\pi i\alpha N}|=1$, 于是

$$I_2<\max|T_\alpha|\cdot\int_{-\frac{1}{\tau}}^{1-\frac{1}{\tau}}|T_\alpha|^2 d\alpha, \tag{62}$$

这里 $\max|T_\alpha|$ 表示当 α 属于辅助区间时 $|T_\alpha|$ 的最大值 $\Big($把因子 $\max|T_\alpha|$ 旁边的积分推广到整个区间 $-\frac{1}{\tau}\leqslant\alpha\leqslant1-\frac{1}{\tau}$, 我们就加强了不等式$\Big)$.

但是复数绝对值的平方等于此数和它的共轭复数的乘积，因此

$$|T_\alpha|^2=T_\alpha\cdot\overline{T}_\alpha,$$

这里由(36)有

$$\overline{T}_\alpha=\sum_{p<N}e^{-2\pi i\alpha p},$$

因为 $e^{-2\pi i\alpha p}=\cos2\pi\alpha p-i\sin2\pi\alpha p$. 所以不等式(62)可以写成

$$|I_2|<\max|T_\alpha|\cdot\int_{-\frac{1}{\tau}}^{1-\frac{1}{\tau}}\sum_{p<N}e^{2\pi i\alpha p}\sum_{p_1<N}e^{-2\pi i\alpha p_1}d\alpha$$

的样子，或写成

$$|I_2|<\max|T_\alpha|\cdot\int_{-\frac{1}{\tau}}^{1-\frac{1}{\tau}}\sum_{p<N}\sum_{p_1<N}e^{2\pi i\alpha(p-p_1)}d\alpha \tag{63}$$

的样子. 但是根据本节开始时所谈到的，在不等式(63)中的积分表示方程式 $p-p_1=0$ 之不大于 N 的素数解 p, p_1 的数目 U, 或者粗略些说，就是不大于 N 的素数的个数 $\Pi(N)$. 根据车比雪夫的结果(12), 有

$$\Pi(N)<B\cdot\frac{N}{\ln N},$$

这里 B 是常数. 这样一来,

$$|I_2|<B\cdot\frac{N}{\ln N}\cdot\max|T_\alpha|, \tag{64}$$

这里再说一遍，$\max|T_\alpha|$ 表示在辅助区间上 $|T_\alpha|$ 的最大值．由于(58)和(59)，对古德巴赫-维诺格拉朵夫定理的证明而言，剩下的是要指出 $\max|T_\alpha|$ 的阶小于 $\frac{N}{(\ln N)^3}$；然而这个事实的建立有着最大的困难，并且是我们所研究的定理的证明中的中心环节．

属于辅助积分中一个积分的每一个 α 表示成 $\alpha=\frac{a}{q}+z$ 的样子，这里 $h<q\leqslant\tau$ 并且 $|z|\leqslant\frac{1}{q\tau}$．这样一来，问题就在于在规定的假设之下来估计三角和

$$T_\alpha=\sum_{p\leqslant N}e^{2\pi i\left(\frac{a}{q}+z\right)p}$$

的绝对值．特别是维诺格拉朵夫确定了

$$\lim_{N\to\infty}\frac{\max T_\alpha}{\dfrac{N}{(\ln N)^3}}=0; \tag{65}$$

在这里，他用了一个由他指出的与我们已经知道的函数 $\mu(n)$ 有关的非常重要的恒等式．

可惜这里不可能给出等式(65)的证明；希望知道这个证明的读者可去研究维诺格拉朵夫的"数论中的三角和方法"一书的第十章．

正如已经指出的那样，从(65)和(64)就得出

$$\lim_{N\to\infty}\frac{I_2(N)}{I_1(N)}=0.$$

这样一来，由于(40)，(58)和(59)就有

$$I(N)=\frac{N^2}{2(\ln N)^3}(S(N)+\gamma(N)), \tag{66}$$

这里 $\lim\limits_{N\to\infty}\gamma(N)=0$，

$S(N)$ 的意义如(60)，并且由(61)，$S(N)>0.6$．定理的证明到这里就结束了．

§5. 整数分解为二平方之和．整复数

在相当大的程度内，素数研究的重要性是由于它们在数之规律性的大部分理论中显然地起着主要作用：常常有这样的问题，初看起来与可除性理论相差很远，在更仔细地研究之后却发现与素数理论有着紧密的联系．我们用下面的例子来说明．

数论中有一个问题是：什么样的自然数可以分解成两个整数(不必异于零)平方之和的问题．

在由可以表为二平方和的数作成的序列中，立刻看不出什么规律性．例如从1到50的数列中，1, 2, 4, 5, 8, 9, 10, 13, 16, 18, 20, 25, 26, 29, 32, 34, 36, 37, 40, 41, 45, 49, 50是二平方和的数，这个序列是相当奇怪的．

十七世纪的法国数学家费尔马曾经指出，全部问题就在于我们提出来的数以何种形式分解为素因数，这本身就是与素数理论有关的问题．

除$p=2$外，素数都是奇数，所以当用4去除时，所得的余数或者是1(形状是$4n+1$的素数)，或者是3(形状是$4n+3$的素数)．

1．素数p当且仅当$p=4n+1$时是二平方和．

形如$4n+3$的数不能是二平方和之证明几乎是显然的：实际上，二偶数的平方和被4除得尽，二奇数的平方和当用4除时余数为2，而一个偶数平方与一个奇数平方的和余数则为1．

对于素数我们先做一点说明，即要证明，如果p是素数则$(p-1)!+1$被p除得尽．p除不尽的数被p除时所得余数为1, 2, 3, …, $p-1$中之一数．我们取整数r, $1\leqslant r\leqslant p-1$，并用$r$去乘1, 2, …, $p-1$；用p去除这样作出来的积，不难证明，得到的余数仍旧全是那些数，但是一般说来次序不同了．特别，余数中有一个是1，就是说，对于每一个r可以找到这样的r_1，使得$r\cdot r_1=1+kp$．我们现证明，仅当$r=1$或$r=p-1$时才有$r=r_1$．事实上，如果$r^2=1+kp$，则$(r+1)(r-1)$被p除得尽；对于数$1\leqslant r\leqslant$

$p-1$，这仅当 $r=1$ 和 $r=p-1$ 时才可能. 我们来求用 p 去除 $(p-1)!=1\cdot2\cdots(p-1)$ 时的余数. 在这个乘积中除了 1 和 $p-1$ 之外，对于每一个因数 r 可以找到与它自己不同的一个 r_1，使得 rr_1 有余数 1. 所以 $(p-1)!$ 被 p 除得的余数就如同全部这些数就是两个因数 1 与 $(p-1)$ 一样，就是说余数为 $p-1$. 这样一来，$(p-1)!+1$ 就被 p 除得尽.

现在令 $p=4n+1$. 我们记

$$(p-1)!+1=\left\{1\cdot2\cdots\frac{p-1}{2}\right\}\times\left\{\left(p-\frac{p-1}{2}\right)\cdots(p-2)(p-1)\right\}-1.$$

在第二个大括号内的式子被 p 除时余数为 $(-1)^{\frac{p-1}{2}}\left(\frac{p-1}{2}\right)!$，但是 $\frac{p-1}{2}=2n$ 是偶数. 所以我们所讨论的情形 $\left(\frac{p-1}{2}\right)!^2+1$ 也能被 p 除得尽. 以 p 除 $\left(\frac{p-1}{2}\right)!$ 所得的余数记作 A. 显然，A^2+1 被 p 除得尽.

我们研究式子 $x-Ay$，其中 x 和 y 互相无关地跑过数零，1，…，$[\sqrt{p}]$（$[x]$ 是不大于 x 的最大整数）. 这样我们就得到 $([\sqrt{p}]+1)^2\geqslant p+1$ 个（相同或不相同）数值 $x-Ay$. 因为被 p 除时不同的余数只可能有 p 个（0，1，2，…，$p-1$），而数值的个数却有 $p+1$ 个，于是就有两对不同的 (x_1, y_1) 和 (x_2, y_2) 使得 x_1-Ay_1 与 x_2-Ay_2 被 p 除时有相同的余数，也就是 $(x_1-x_2)-A(y_1-y_2)$ 被 p 除得尽. 我们假定：$x_0=x_1-x_2$，$y_0=y_1-y_2$. 显然 $|x_0|<\sqrt{p}$，$|y_0|<\sqrt{p}$. 既然 A^2+1 被 p 除得尽，故 $y_0^2(A^2+1)=(Ay_0)^2+y_0^2$ 被 p 除得尽；但是因为 x_0-Ay_0 被 p 除得尽，故 p 除得尽 $x_0^2-(Ay_0)^2=(x-Ay_0)(x+Ay_0)$. 所以等于 $(x_0^2-(Ay_0)^2+(Ay_0)^2+y_0^2)$ 的数 $x_0^2+y_0^2$ 被 p 除得尽. 但是 $|x_0|<\sqrt{p}$，$|y_0|<\sqrt{p}$. 由此我们得出结论，或者 $x_0^2+y_0^2$ 是零，或者 $x_0^2+y^0=p$. 第一种情形不可能，因为 (x_1, y_1)，(x_2, y_2) 是不同的两对. 因此，形如 $4n+1$ 的素数可以表成二平方和的形式.

2. 我们现来讨论将任意整数分解为二平方和的问题. 容易验证恒等式

$$(a^2+b^2)(c^2+d^2)=(ac-bd)^2+(ad+bc)^2.$$

这个恒等式表明,两个是二平方和的整数的乘积仍是二平方和. 由此得出, 2 以及形如 $4n+1$ 的素数的任何次方的乘积仍是二平方和. 因为二平方和乘上一个平方仍得到二平方和,于是形如 $4n+3$ 的素因子出现偶数次方的任何数是二平方和.

3. 我们证明, 如在一数中形如 $4n+3$ 的素数出现奇数次方, 则该数不能分解为二平方和. 这样, 以前所提出的问题即可完全解决.

我们现在来研究形如 $a+bi$ 的复数, 这里 a 和 b 是普通的整数. 这种复数叫做整复数. 如果整数 N 是二平方和 $N=a^2+b^2$, 则 $N=(a+bi)(a-bi)=\alpha\bar{\alpha}$ (我们用 $\bar{\alpha}$ 表示 α 的共轭复数), 于是 N 在整复数范围内可以分解为复共轭因子的乘积.

在整复数范围内可以建立与普通整数范围内可除性理论完全类似的可除性理论. 如果 $\frac{\alpha}{\beta}$ 仍是整复数, 我们就说整复数 α 被整复数 β 除得尽. 只存在四个除得尽 1 的整复数,这就是 $1, -1, i, -i$. 如果整复数 α 除了 $1, -1, i, -i, \alpha, -\alpha, \alpha i, -\alpha i$ 之外没有别的因子, 我们就说 α 是素数. 在第一段中所解决的问题现在有了新的意义: 那里说明,形如 $4n+1$ 的素数和 2 在整复数范围内不再是素数. 容易证明,形如 $4n+3$ 的素数还是素数. 实际上如果 $p=\alpha\beta$, 则 $p=\bar{\alpha}\bar{\beta}$, $p^2=\alpha\bar{\alpha}\beta\bar{\beta}$. 但 $\alpha\bar{\alpha}$ 和 $\beta\bar{\beta}$ 是普通的正整数; $p\neq\alpha\bar{\alpha}$, 因为形如 $4n+3$ 的素数不是二平方和. 可见 $\alpha\bar{\alpha}=1$ 就是可能或者是 ±1, 或者是 $\pm i$, 就是说除了明显的因子之外 p 没有别的因子.

对于整复数而言, 分成素因子的唯一分解定理是正确的. 如果撇开因子的次序及它们与数 $1, -1, i, -i$ 的结合,当然唯一性就得到了.

令 N 是平方和, $N=\alpha\bar{\alpha}$. 令 p 是形如 $4n+3$ 的素数. 我们来

计算在数 N 中 p 有几次方，因为 p 在整复数范围内还是素数，计算在 α 中和在 $\bar{\alpha}$ 中各有 p 的几次方就够了. 但这些方次是相等的，所以 p 在 N 中一定有偶次方，这就是所要证明的.

内容丰富的可除性理论不仅在有理整数范围内成立（正如我们看到的那样，它在整复数范围内成立）这个事实的发现大大扩大了十九世纪数学家们的眼界. 对这些概念的研究要求在数学中建立新的一般的概念，例如环的概念和理想的概念. 现在这些概念的意义已远远超出了数论的范围.

文 献

通 俗 读 物

А. Г. Постников и Н. П. Романов, Упрощение элементарного доказательства А. Сельберга асимптотического закона распределения простых чисел. "Успехи матем. наук", 10, № 4, 1955.

文中有渐近规律之证明的通俗的叙述，可作为史尼雷尔曼的小册子的补充。

Л. Г. Шнирельман, Простые числа, Гостехиздат, 1940.

教 科 书

И. В. Арнольд, Теоретическая арифметика, Изд. 2-е, Учпедгиз, 1939.

И. М. 维诺格拉朵夫，数论基础，高等教育出版社，1956年版.

П. Л. Чебышев, теория сравнений. Полное собрание, сочинений т I. Изд-во АН СССР, 1944.

专 题 论 文

И. М. 维诺格拉朵夫，数论中的三角和方法，В. А. 斯捷克洛夫数学研究所报告书，第23期，1947，数学进展，第一卷第一期，科学出版社，1955年.

М. Интам. Распределение простых чисел ГОНТИ, 1936.

华罗庚，堆垒素数论，В. А. 斯捷克洛夫数学研究所报告书，第22期，1947（中文本）.

许孔时 译

越民义 王 元 校

第十一章 概 率 论

§1. 概率规律性

自然科学所确立的最简单的规律性，可以归结为指出我们所感兴趣的某些事件必然发生或必然不发生的条件，换句话说，这些条件可以表成以下的两种概型中的一种：

1) 如果条件组 S 实现，则事件 A 必然发生；

2) 如果条件组 S 实现，则事件 A 必然不发生.

在前一种情形下，事件 A 对于条件组 S 而言可以称为"确切"事件或"必然"事件，而在后一种情形下，事件 A 对于条件组 S 而言可以称为"不可能"事件，例如，在大气压力及温度 $t(0°<t<100°)$ 的情形下(条件组 S)，水一定处于液体状态(必然事件 A_1)，而不能处于气体状态或固体状态(不可能事件 A_2 与 A_3).

如果当条件组 S 实现时，事件 A 有时发生有时不发生，则称 A 为对于条件组 S 而言的随机事件. 于是就产生了这样的问题：事件 A 的随机性是否意味着事件 A 与条件组 S 之间的联系没有任何规律性呢？例如假定我们知道某厂所生产的某种类型的灯泡(条件组 S)有时可以燃点 2,000 小时以上(事件 A)，而有时则不到这个时间就已经烧坏了. 虽然如此，但是我们是否可以就用检查灯泡燃点 2,000 小时的能力的实验来鉴定这个厂的产品质量呢？或者我们是否应该只限于指出实际上所有灯泡可以连续燃点的时数(例如说，500 小时)以及实际上足以使所有灯泡尽皆烧坏的时数(例如说，10,000 小时)呢？显然，仅只用不等式 $500\leqslant t\leqslant 10,000$ 来说明灯泡燃点时数是不能使消费者满意的. 如果能够告诉消费者，差不多在 80% 的情形下，灯泡燃点的时数不少于 2,000 小时，则消费者将会获得充分得多的了解. 灯泡质量的更

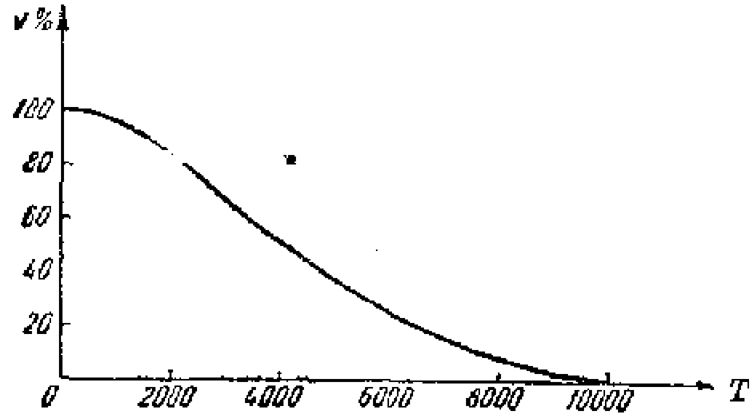


图 1

加充分的鉴定是对于任何时间 T 给出燃点时数不少于 T 的那些灯泡的百分比 $\nu(T)$，例如可以用图 1 所示的图线这种形式来给出。

曲线 $\nu(T)$ 实际上可以由对于充分多的一部分灯泡(100—200个)作试验而得出。自然，欲使这样得出的曲线有实际价值，必须使它正确地反映出在原料的一定质量及工厂中所规定的生产工艺规程的条件下生产出来的全部(而非只是所试验的这一部分)灯泡的真实规律，换句话说，必须在对于任何一组(在这些一般条件下生产出来的)灯泡来作试验时恒能得出相似的结果(即得出与试验第一组灯泡时所得的曲线相差不多的曲线 $\nu(T)$)。这就是说，在所试验的那些部分中，各曲线 $\nu(T)$ 所表现出的统计规律性仅只是灯泡燃点时数与原料质量及生产工艺规程之间的概率规律性的反映。

这种概率规律性可以借助于函数 $\mathbf{P}(T)$ 来给出，此处 $\mathbf{P}(T)$ 表示(在给定条件之下生产出来的)个别灯泡燃点时数不少于 T 小时的概率。

我们断言事件 A 在条件 S 之下有确定的概率

$$\mathbf{P}(A/S)=p,$$

这意思就是说，在各个充分长的试验(即实现条件组 S)序列中所得到的事件 A 实现的频率

$$v_r = \frac{\mu_r}{n_r}$$

大致相同并且接近于 p (此处 n_r 是第 r 个试验序列中的试验次数, μ_r 是在这 n_r 次试验中事件 A 出现的次数).

我们仅只是假定有这种常数 $p=\mathbf{P}(A/S)$ 存在(客观上系由条件组 S 与事件 A 之间的关系的特性来决定), 使得"一般说来"当试验次数 n 愈大时频率 ν 就愈与它靠近; 这种假定可以在很广泛的一类现象中得到良好的证实. 我们自然就称这一类现象为**概率随机现象**[1].

以上所考察的例系属于大量生产的概率规律性范畴. 这种规律性的现实性是不容置疑的. 大量生产中统计抽样检查的许多最重要的实用方法就正是以这种规律性为基础的. 从概率规律性的形成方式方面看来, 对于射击论有根本意义的炮弹分布的概率规律系属于与此相近的范畴. 因为从历史上说这是概率论方法应用到技术问题上去的最早事例之一, 所以在下文中我们还要再讲到射击论中若干最简单的问题.

上文中我们说到当试验次数 n 很大时, 频率 ν 与概率 p "靠近", 这些话多少有一些模糊; 我们未曾说明对于各种 ν 值, 差数 $\nu-p$ 将小至何种程度. 我们将在 § 3 中给出 ν 与 p 的靠近程度的数量估计. 有意思的是可以指出, 在这个问题中完全排除某种不确切性是办不到的. 因为当问题确切化时就可发现关于 ν 与 p 相靠近的断言本身只具有概率特性.

§ 2. 初等概率论的公理与基本公式

既然统计规律性的重大作用是无可置疑的, 从而就产生了研究这种规律性的方法的问题. 首先产生的想法即是能否以纯粹经验的、试验的方法来研究这种规律性. 由于概率规律性呈现在大量过程里, 因之要发现这种规律似乎自然就需进行大量的试验.

1) 有时称为斯笃哈斯谛现象.

然而这种概念却仅只是部分地合于实情．当我们用试验方法建立了某一些概率规律之后，我们就可以借助于若干一般性的假定，利用逻辑方法或计算方法，从这些规律推得新的概率规律．在说明这应该如何来作之前，我们需先列举出概率论的若干基本定义与公式．

因为我们是把概率了解为频率 $\nu=\frac{m}{n}$（此处 $0\leqslant m\leqslant n$，因之 $0\leqslant\nu\leqslant 1$）的正常数值，所以我们自然规定任何事件 A 的概率 $\mathbf{P}(A)$ 是介于 0 与 1 之间的[1]，

$$0\leqslant\mathbf{P}(A)\leqslant 1. \qquad (1)$$

如果两个事件（当条件组 S 实现时）不可能同时出现，则称之为**不相容的**．例如掷骰子时出偶数点与出三点是不相容事件．如果把事件 A_1 与 A_2 中至少有一件出现算作是事件 A，则我们称事件 A 为事件 A_1 与 A_2 的**联合**．例如掷骰子时若把 1，2 或 3 点的出现作为事件 A，1 或 2 点的出现作为事件 A_1，2 或 3 点的出现作为事件 A_2，则事件 A 是事件 A_1 与 A_2 的联合．易见，若把两个不相容事件 A_1，A_2 以及它们的联合 $A=A_1\mathbf{U}A_2$ 的出现次数记作 m_1，m_2 与 m，则恒有等式 $m=m_1+m_2$，由此可见，对于相应的频率恒有等式 $\nu=\nu_1+\nu_2$．

由此就很自然地引出下列的概率加法公理

$$\mathbf{P}(A_1\mathbf{U}A_2)=\mathbf{P}(A_1)+\mathbf{P}(A_2). \qquad (2)$$

此处事件 A_1 与 A_2 是不相容的，而 $A_1\mathbf{U}A_2$ 系表示它们的联合．

其次，对于必然事件 U，我们自然取

$$\mathbf{P}(U)=1. \qquad (3)$$

全部概率数学理论都是建立在象(1)，(2)，(3)这样简单的公理的基础上的．由纯数学观点看来，“概率”乃是“事件”的数值函数，这个函数具有某些公理化所固定下来的性质．如果我们不坚持要求事件、事件的联合以及下文中将要定义的事件的兼有这样一些概念本身也必须公理化，则公式 (1)，(2)，(3) 所表达的概率

1) 为了书写简便，现在我们把 $\mathbf{P}(A/S)$ 改写为 $\mathbf{P}(A)$．

性质即是建立所谓的初等概率论的充分的基础. 初学的人对于"事件"与"概率"这些名词最好采取直觉的了解法, 不过知道这样一点也是有益的, 即这些概念的未经完全形式化的实际意义并不影响概率论的公理化了的纯粹数学讲法的纯形式上的清晰性.

设任意给定若干个事件 $A_1, A_2, \cdots, A_s$, 而以 A 表示这些事件中至少有一个实现, 则我们称事件 A 是这些事件的联合. 于是由加法公理容易推知, 对于任意多个彼此互不相容的事件 $A_1, A_2, \cdots, A_s$ 及其联合 A, 我们有

$$\mathbf{P}(A)=\mathbf{P}(A_1)+\mathbf{P}(A_2)+\cdots+\mathbf{P}(A_s)$$

(所谓"概率加法定理").

如果这些事件的联合是必然事件(即每当条件组 S 实现时, 总有某一事件 A_k 实现), 则

$$\mathbf{P}(A_1)+\mathbf{P}(A_2)+\cdots+\mathbf{P}(A_s)=1,$$

此时事件组 $A_1, \cdots, A_s$ 称为完备事件组.

现在我们试来考察两个一般说来彼此相容的事件 A 与 B. 如果用事件 C 表示事件 A 与 B 同时发生, 则事件 C 称为事件 A 与 B 的兼有 $(C=AB)$[1].

例如, 设事件 A 表示当掷一个骰子时出现偶数点, 而事件 B 表示出现的点数能被三整除, 则事件 C 就是出现六点.

假定当把一个试验重复进行到 n (大数)次时, 事件 A 出现了 m 次, 而事件 B 出现了 l 次, 其中有 k 次是和事件 A 同时出现的. 于是比值 $\frac{k}{m}$ 自然可称为事件 B 在事件 A 的条件下的条件频率. 频率 $\frac{k}{m}$, $\frac{m}{n}$ 与 $\frac{k}{n}$ 之间的关系公式

$$\frac{k}{m}=\frac{k}{n}:\frac{m}{n}$$

使我们自然地平行引出如下的定义:

比值 $$\mathbf{P}(B/A)=\frac{\mathbf{P}(AB)}{\mathbf{P}(A)}.$$

1) 类似地, 任意多个事件 $A_1, A_2, \cdots, A_s$ 的兼有 C 表示这些事件同时一起发生.

称为事件 B 在事件 A 的条件下的**条件概率**.

当然,这里我们假定 $\mathbf{P}(A)\neq 0$.

如果事件 A 与 B 在本质上彼此毫无联系,则我们自然可以预料,事件 B 在出现事件 A 的条件下不应比考察所有一般的试验时出现得频繁得多或出现得稀少得多,换句话说,即我们应有近似公式 $\frac{k}{m}\sim\frac{l}{n}$ 或

$$\frac{k}{n}=\frac{k}{m}\frac{m}{n}\sim\frac{l}{n}\frac{m}{n}.$$

在最后的这个近似公式中,$\frac{m}{n}=\nu_A$ 是事件 A 的频率,$\frac{l}{n}=\nu_B$ 是事件 B 的频率,$\frac{k}{n}=\nu_{AB}$ 是事件 A 与 B 的兼有 AB 的频率.

我们看到,这些频率由下列的近似公式

$$\nu_{AB}\sim\nu_A\nu_B$$

相互联系. 因此,对于事件 A, B 与 AB 的概率自然就应该有相应的确切等式

$$\mathbf{P}(AB)=\mathbf{P}(A)\cdot\mathbf{P}(B). \tag{4}$$

等式(4)就作为两个事件 A 与 B 的独立性的定义.

类似地可以定义任意多个事件的独立性. 此外,还可以给出任意多个试验的独立性的定义(这个定义粗略说来可以归结成这样一点,即在这些试验中,一部分试验的结果不影响其余试验的结果)[1].

设有 n 个独立试验,在每一个试验中事件 A 出现的概率皆等于 p. 现在我们试来计算在这 n 个试验中事件 A 恰巧出现 k 次的概率 P_k. 我们用 $\bar{A}$ 表示"事件 A 不出现"这个事件. 显而易见,

---

1) 确切说来,"试验的独立性"的意义如下:把 n 个试验用任何方法划分为两组,设事件 A 表示第一组中所有的试验皆以某些预先给定的结局而完结,而事件 B 表示第二组中所有的试验皆以某些预先给定的结局而完结. 如果对于任意的划分法与任意给定的结局,以上所定义的事件 A 与 B 恒在 (4) 的意义下彼此独立,则这些试验可以称为是(完全)"独立的".

在§4中我们还要再来讨论独立性概念的现实意义.

$$\mathbf{P}(\bar{A})=1-\mathbf{P}(A)=1-p.$$

从试验的独立性的定义不难看出．由事件 A 出现 k 次，不出现 $n-k$ 次所组成的任何一个确定的序列，其概率皆为

$$p^k(1-p)^{n-k}. \tag{5}$$

例如，当 $n=5$ 与 $k=2$ 时，得到 $A\bar{A}A\bar{A}\bar{A}$ 这一串结果的概率为

$$p(1-p)p(1-p)(1-p)=p^2(1-p)^3.$$

根据加法定理，概率 P_k 等于所有由事件 A 出现 k 次，不出现 $n-k$ 次所组成的结果序列的概率的和数；换句话说，根据 (5)，概率 P_k 等于这种结果序列的个数与 $p^k(1-p)^{n-k}$ 的乘积．显而易见，这种序列的个数等于从 n 个里取 k 个的组合数，因为 A 的 k 次出现可以在 n 个试验的序列中取任何位置．

这样，最后我们便得到

$$P_k=C_n^k p^k(1-p)^{n-k} \qquad (k=0,\ 1,\ 2,\ \cdots,\ n) \tag{6}$$

(所谓二项式分布)．

为了要看看以上所引入的定义与公式如何应用，我们试来考察一个有关射击论的例子．

假定为了击破目标，有五发命中就够了．我们要研究的问题是：我们是否可以指望在 40 发射击中得到这所要的五发命中？这个问题的纯经验的解决方法如下：对于一定大小的目标，在一定的距离上进行射击，每批射击 40 发，如此进行许多批 (例如说 200 批)；然后查明在多少批射击中得到了不少于五发的命中．比如说，在 200 批射击中有 195 批射击得到了这样的成绩，则概率 P 近似地等于

$$P=\frac{195}{200}=0.975.$$

依照上述的纯经验的研究方法来作，为了解决这个极特殊的问题就得浪费 8000 发炮弹．所以实际上当然谁也不这么办．我们可以舍此而从研究在(与目标大小无关的)固定射击距离下的炮弹分布入手．事实上，对于平均落点的远近偏差及左右偏差，就各

种大小偏差发生的频率而言，服从图 2 所示的规律. 此处 B 系表示所谓的或然偏差. 一般说来，对于远近偏差的或然偏差和对于左右偏差的或然偏差是不同的，而且除此之外，或然偏差还随着射击距离的增长而增长. 对于各种距离，各种类型的大炮与炮弹，其或然偏差可借助于炮兵射击场的试射而以经验方法求得. 作过了这一步以后，所有象以上所提出的那种类型的特殊问题，都可以用计算的方法来解决.

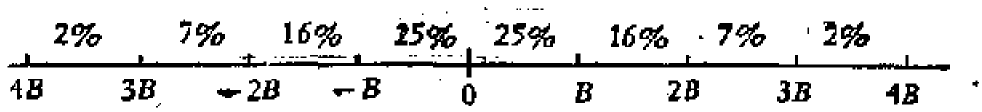


图 2

为了简单起见，假定我们的目标是长方形的，其一边与射击线平行，长为远近或然偏差的两倍，另一边与射击线垂直，长为左右或然偏差的两倍. 其次，我们假定目标很好地被试射了，并且平均弹道穿过目标的中心(图 3).

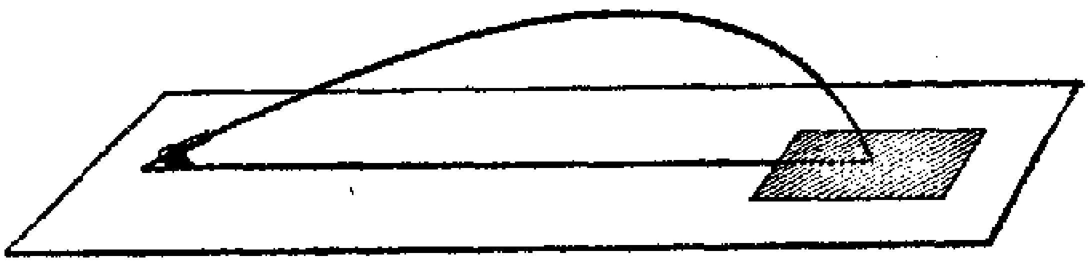


图 3

我们再假定左右偏差与远近偏差彼此独立[1]. 于是为了要使某发炮弹命中目标，必须而且只须其远近偏差与左右偏差不超过相应的或然偏差. 根据图 2，这两件事件中的每一件皆大致对于 50% 的所发射的炮弹可以实现，换句话说，每一事件实现的概率皆为 1/2. 这两件事件的兼有大致对于 25% 的所发射的炮弹可以实现，换句话说，个别炮弹命中目标的概率等于

$$p=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4},$$

1) 这种独立性的假定可由试验证实.

而个别炮弹不命中目标的概率等于

$$q=1-p=1-\frac{1}{4}=\frac{3}{4}.$$

假定不同的各发炮弹的命中是独立事件，利用二项式公式(6)，即可求得在40发炮弹中恰有 k 次命中的概率等于

$$P_k=C_{40}^k p^k q^{40-k}=\frac{40\cdot 39\cdots(40-k)}{1\cdot 2\cdots k}\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)^{40-k}.$$

我们所要求的不少于五发命中的概率现在可以表如以下的公式：$P=\sum_{k=5}^{40}P_k$. 但是从少于五发命中的概率 $Q=\sum_{k=0}^{4}P_k$ 出发，按照公式 $P=1-Q$ 来计算 P，则比较简单.

我们可以算出，

$$P_0=\left(\frac{3}{4}\right)^{40}\sim 0.00001,$$

$$P_1=40\left(\frac{3}{4}\right)^{39}\frac{1}{4}\sim 0.00013,$$

$$P_2=\frac{40\cdot 39}{2}\left(\frac{3}{4}\right)^{38}\left(\frac{1}{4}\right)^2\sim 0.00087,$$

$$P_3=\frac{40\cdot 39\cdot 38}{2\cdot 3}\left(\frac{3}{4}\right)^{37}\left(\frac{1}{4}\right)^3\sim 0.0037,$$

$$P_4=\frac{40\cdot 39\cdot 38\cdot 37}{2\cdot 3\cdot 4}\left(\frac{3}{4}\right)^{36}\left(\frac{1}{4}\right)^4\sim 0.0113,$$

因此，$Q=0.016$, $P=0.984$.

这里所得到的概率 P，比起在确定能够保证完成上述任务的炮弹数时射击论中通常认为足够的概率，还要更接近于1一些. 通常认为指出能以概率0.95保证完成上述任务的炮弹数目就可以了.

以上我们所考察的例子虽然多少是概型化了的，但是却充分使人信服地表明了概率计算的重要性. 由试验确定了或然偏差对于射击距离的依赖关系之后(这只要在射击场进行不很多次的射击就够了)，再通过一些不甚复杂的计算，我们就可以得出各种各样问题的答案了. 在所有由于大量随机因素的共同作用而产生统

计规律性的场合，情形亦皆与此类似．对于大量观察直接加工整理的结果，只能显示出这些统计规律性中的最简单的一些，换言之，即只能求得某些基本概率．然后，利用概率论的规律，我们就可以从这些基本概率出发，计算出更复杂的现象的概率，并且从而推断我们所考察的那些复杂现象的统计规律性．

有时不用积累大量统计材料，问题也完全能获得解决，因为基本概率可能由充分令人信服的对称性考虑而得以确定．例如，自古以来我们就能够肯定，骰子（亦即由均匀材料所雕成的正方体）从充分高的高度掷下来时，每一面朝上的概率都是 1/6，而我们作出这一项断言无疑是先于十分系统地去积累为要证实这一断言所需的充分的观察材料．这种类型的系统试验主要是十八至二十世纪的概率论教材作者们作的，彼时概率论已经是一门经过研究的科学了．这个核验的结果是令人满意的，但是把这种作法推广到新的类似的情形中去，就未必有什么意思了．例如，我们知道，谁也没有对于抛掷由均匀材料所雕成的规则十二面体一事进行过充分大量的试验．不过毫无疑问，倘若真来进行试验，比如说抛掷 12,000 次，则十二面体任何一面朝上的次数皆约略为一千次．

由对于对称性或均匀性的考虑而推求基本概率的这种方法，对于许多重大的科学问题也是具有重要意义的；例如在所有关于毫无秩序地运动着的气体分子的碰撞或接近的问题上，情形就是这样，对于所有关于银河中的星的碰撞或接近的问题，这种方法也同样有效．当然，在象这种更精微的场合，对于所作的假定最好能验证一下，即使是通过把从所作的假定而推得的结论与试验相比较这种办法来间接验证也好．

§3. 大数定律与极限定理

关于在“长的”试验序列中事件出现的频率“接近”于事件的概率这一断言，非常自然地需要以数量来确切说明．应该弄清楚这一问题的某种微妙性．在概率论中最典型的场合下，对于任意长

的试验序列，频率的两个极端值

$$\frac{\mu}{n}=\frac{n}{n}=1 \quad 与 \quad \frac{\mu}{n}=\frac{0}{n}=0$$

在理论上总还是可能的. 因此，试验次数 n 无论怎样大，我们始终不能完全断定我们有这样的不等式：

$$\left|\frac{\mu}{n}-p\right|<\frac{1}{10}.$$

例如，假定事件 A 是掷骰子时出现六点，则当投掷 n 次时，我们得到的总是六点的概率为 $\left(\frac{1}{6}\right)^n>0$，换句话说，六点出现的频率等于 1 的概率为 $\left(\frac{1}{6}\right)^n$，而六点一次也不出现，换句话说，六点出现的频率等于零的概率为 $\left(1-\frac{1}{6}\right)^n>0$.

在所有类似的问题中，对于频率与概率的靠近程度的任何非显然的估值都不能完全确定地成立，而只是以某一小于 1 的概率成立. 例如，可以证明，在事件出现具有固定概率 p 的独立试验[1]的情形下，下列不等式

$$\left|\frac{\mu}{n}-p\right|<0.02 \tag{7}$$

对于 $n=10,000$（及任意的 p）成立的概率为

$$P>0.9999. \tag{8}$$

在这里我们首先应着重指出，在以上的叙述中，频率 $\frac{\mu}{n}$ 与概率 p 的靠近程度的数值估计(7)是和所引入的新的概率 P 联结在一起的，而决不能孤立地来看待它.

估值式(8)的实际意义是这样的：假定进行 N 批试验，每一批试验由 n 次试验组成，然后计算出使不等式(7)成立的批数 M，那么当 N 充分大时，我们将近似地有

$$\frac{M}{N}\approx P>0.9999. \tag{9}$$

1) 参看 270 页注解. 估值(8)的证明见 273 页.

但是如果我们想要在 $\frac{M}{N}$ 与 P 的靠近程度方面以及断言这种靠近程度能够成立的确实性方面，来使关系式 (9) 精确化，那么我们就必须象在 $\frac{\mu}{n}$ 与 p 的靠近程度的问题上所作的那样，重新来作类似的考虑。如若高兴的话，这种考虑可以重复任意多次，但是十分明显，这并不能使我们获得根本的解脱，到了最后一步我们仍然不得不对于概率这一个词采取原来的粗糙了解。

不要以为这一类的困难乃是概率论的某种固有的性质。在对现实现象进行数学研究时，我们永远是要把它们概型化。实际现象的演变与理论概型之间的偏差也可以加以数学研究。但是为此这些偏差本身就需要纳入一定的概型，而在使用这一定的概型时就已经不必再对这些偏差与这一定的概型之间的偏差进行形式数学分析了。

然而我们指出，当我们把估值式[1]

$$\mathbf{P}\left\{\left|\frac{\mu}{n}-p\right|<0.02\right\}>0.9999 \tag{10}$$

实际应用到由 n 次试验所组成的单独一批试验上去的时候，我们依据的乃是某种对称性的考虑：不等式 (10) 是说，当批数 N 很大的时候，关系式 (7) 将在不少于 99.99% 的场合下成立；如果我们有根据认为我们所考察的这一批试验在其他各批试验中占有的只是一个普通的，毫无特点的位置，那么自然我们就有很大的把握来期待不等式 (7) 对于我们所考察的这特别的一批试验成立。

通常在各个不同的实际情形下所忽略的概率是各不相同的。在前文中已经指出，当预计为要保证完成任务所需的炮弹数量时，一般认为取定能以概率 0.95 完成任务所需的炮弹数量就可以了，换句话说，忽略了不超过 0.05 的概率。这是由于倘若我们在计算时只准许忽略例如说小于 0.01 的概率，那么就会使得炮弹数量大大增加，这也就是说，实际上在许多情形下将会推论出在所能容许的短促时间之内完成任务是不可能的，或者说具有真正威力的是

1) 对于不等式 (7) 成立的概率的估值式 (8) 可以写成 (10) 的样式。

所使用的炮弹储备.

有时在科学研究中也只是使用以不计 0.05 以内的概率为前提的统计方法. 但仅只是当收集更广泛的材料非常困难时才可以这样作. 作为运用这种方法的例, 我们试来考察以下的问题. 假定在一定的条件之下, 治疗某种疾病的通用药剂在 50% 的情形下, 换句话说, 以概率 0.5 给出良好结果. 现在我们提出了一种新的药剂, 并且为了要证实它确比旧有的药剂优越, 我们从病人(病情的严重程度是这样的, 即旧有药剂在这种情形下有 50% 的疗效) 里面毫无偏心地选择十个人来试用这种新的药剂. 此时如果在十个人中有不少于八个人皆得到了良好的结果, 则新的药剂的优越性可以认为是被证实了. 不难算出, 在问题的这种解法中得到错误结论(所谓错误结论即当新的药剂与旧的药剂不相上下甚至还更差一些的时候, 我们断言新的药剂的优越性已被证实)的概率恰在 0.05 左右, 这一概率被忽略不计. 实际上, 如果十个试验中每一个试验得到良好结果的概率皆为 p, 则在十个试验中共得到 10 个, 9 个或 8 个良好结果的概率分别等于

$$P_{10}=p^{10},\ P_9=10p^9(1-p),\ P_8=45p^8(1-p)^2.$$

当 $p=\frac{1}{2}$ 时, 则我们加起来便得到

$$P=P_{10}+P_9+P_8=\frac{56}{1024}\sim 0.05.$$

这样一来, 假若新的药剂真的与旧有的药剂不相上下, 则我们错误地断言新的药剂优于旧有的药剂的概率在 0.05 左右. 若是我们要把这个概率降低到 0.01 左右, 同时不增加试验的数目 $n=10$, 那么我们就必须规定, 只有当新的药剂在十次试验中不少于九次得到良好结果时, 新的药剂的优越性才能认为是被证实了. 如果这个要求对于新的药剂的拥护者说来是过于严格了, 那么就必须使试验的次数 n 大大地超过 10. 例如, 假定当 $n=100$ 时我们规定, 若是 $\mu>65$ 则新的药剂的优越性就算是被证实了, 那么错误的概率仅为 $P\approx 0.0015$.

假如说 0.05 这种大小对于一些重要的科学研究还是很不够的，那么象 0.001 或 0.003 这样大小的错误概率，在绝大多数情形下，甚至在象天文观测结果的整理加工这样的典型的和精密的研究中，一般也都是忽略不计的．不过有时候以概率规律性为基础的科学论断也具有极高的准确性(这就是说，在作出这种论断时只忽略极小的概率)．关于这一点下文中还要谈到．

在 n 个独立试验中(设每一试验得到正面结果的概率皆为 p)，恰恰得到 m 个正面结果的概率 P_m 可由二项式公式(6)来表出：

$$P_m=C_n^m p^m(1-p)^{n-m};$$

在我们以前所考察的几个例中我们已经屡次使用了这一公式的各种特殊情形．现在我们再利用这个公式来研究我们在本节开头处所提出的关于概率

$$p=\mathbf{P}\left\{\left|\frac{\mu}{n}-p\right|<\varepsilon\right\} \tag{11}$$

的问题，此处 μ 表示实际得到的正面结果的次数[1]．显而易见，这个概率可以表成以满足下列不等式

$$\left|\frac{m}{n}-p\right|<\varepsilon \tag{12}$$

的那些 m 值为下标的那些 P_m 的和数，即

$$P=\sum_{m=m_1}^{m_2} P_m, \tag{13}$$

其中 m_1 表示满足不等式(12)的那些 m 值中的最小者，而 m_2 表示其中的最大者．

当 n 稍大时，公式(13)对于直接计算来说是不大合适的．因此莫哇佛尔对于 $p=\frac{1}{2}$ 以及拉普拉斯对于任意的 p 所发现的渐近公式就有很大的意义，渐近公式使得我们对于大的 n 值可以很简捷地求出概率 P_m 的值并且研究它的性质．这个渐近公式的形状如下：

---

1) μ 以概率 P_m 取值 $m(=0,1,\cdots,n)$，亦即

$$\mathbf{P}(\mu=m)=P_m.$$

$$P_m \sim \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(m-np)^2}{2np(1-p)}}, \tag{14}$$

如果 p 不是非常接近于零或 1, 则当 n 等于 100 左右的时候, 这个公式就已经十分地精确了. 如果我们令

$$t = \frac{m-np}{\sqrt{np(1-p)}}, \tag{15}$$

则公式(14)可以改写成

$$P_m \sim \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{t^2}{2}}. \tag{16}$$

由(13)及(16)可以推出概率(11)的如下的近似表达式:

$$P \sim \frac{1}{\sqrt{2\pi}} \int_{-T}^{T} e^{-\frac{t^2}{2}} dt = F(T), \tag{17}$$

其中

$$T = \varepsilon \sqrt{\frac{n}{p(1-p)}}. \tag{18}$$

对于固定的不等于零或 1 的 p, (17)式左右两侧的差当 $n \to \infty$ 时对 ε 而言一致地趋于零. 对于函数 $F(T)$ 我们有许多详尽的表. 在此我们只摘引出如下的值:

T	1	2	3	4
F	0.68269	0.95450	0.99730	0.99993

当 $T \to \infty$ 时, 函数 $F(T)$ 的值趋于 1.

现在我们试利用公式 (17) 来对于 $n=10,000$ 估计以下的概率:

$$\mathbf{P}\left\{\left|\frac{\mu}{n}-p\right| < 0.02\right\}.$$

因为 $T = \frac{2}{\sqrt{p(1-p)}}$, 所以

$$P \approx F\left(\frac{2}{\sqrt{p(1-p)}}\right).$$

由于函数 $F(T)$ 随着 T 的增加而单调递增, 因之为了要得出不依赖于 p 的 P 的最低估值, 应取所有可能的 T 值 (对应于不同

的 p) 中的最小值. 当 $p=\frac{1}{2}$ 时 T 达到最小值, 并且它等于 4. 所以我们近似地有

$$P \geqslant F(4)=0.99993. \tag{19}$$

在不等式(19)中没有考虑到由于公式(17)的近似性而产生的误差. 对于由此而产生的误差进行估值之后, 我们可以断言, 在任何情形下恒有 $P>0.9999$.

从我们所考察的关于公式(17)的应用的这个例自然就引起了公式(17)的余项的估值问题, 应该指出, 在概率论理论著作中给出的这种估值很久以来都是停留在很不能令人满意的地步. 因此对于不很大的 n 或者对于很靠近 0 或 1 的概率 p (这种概率往往偏具有特别重大的意义), 应用公式(17)及与之类似的公式来作计算的时候, 时常仅只是对于有限个模型来核验这种结果, 以此为其依据, 而并未对可能有的错误建立精确的估值. 除此之外, 更详尽的研究表明, 在许多具有实际重要性的情形中, 以上所引入的渐近公式不仅是需要对它的余项进行估值而且公式本身还需要精确化(因为如果不把公式作得更精确一些, 余项就过于大了). 在这两方面最完全的结果系属于伯恩斯坦.

公式(11), (17)与(18)可以改写成以下的形状:

$$\mathbf{P}\left\{\left|\frac{\mu}{n}-p\right|<t \sqrt{\frac{p(1-p)}{n}}\right\} \sim F(t). \tag{20}$$

公式(20)右侧不包含 n, 对于充分大的 t 它任意接近于 1, 换句话说, 它任意接近于表示完全确切性的概率值. 由此可见, 一般说来, 频率 $\frac{\mu}{n}$ 与概率 p 之间的偏差的级为 $\frac{1}{\sqrt{n}}$. 概率规律作用的精确性与观察次数的平方根之间的这种比例关系也是许多其他问题所共有的现象. 有时简单通俗一些, 我们甚至就说"n 的平方根规律"是概率论的基本规律. 这种思想的完全明确系归功于伟大的俄罗斯数学家车比雪夫, 由于他首先系统地使用了把各种不同的概率问题归结到计算"随机变量"的和数与算术平均值的"数学期望"与"离差"的方法.

如果在某一个条件组 S 之下，一个变量能以各种确定的概率取各种不同的值，则称此变量为随机变量．对于我们说来，只要来考察仅能取有穷多个不同的值的随机变量就够了．为了要说明这种随机变量 ξ 的所谓的概率分布，只须指出它的所有可能值 x_1, x_2, $\cdots$, x_s 以及概率

$$P_r=\mathbf{P}\{\xi=x_r\}.$$

这些概率按照变量 ξ 的所有各种可能的值求和时，其和数永远为 1，

$$\sum_{r=1}^{s}P_r=1.$$

我们以上考察过的 n 个试验中的正面结果的数目 μ 可以作为随机变量的例．

下列表达式

$$\mathbf{M}(\xi)=\sum_{r=1}^{s}P_r x_r,$$

称为变量 ξ 的**数学期望**，而偏差 $\xi-\mathbf{M}(\xi)$ 的平方的数学期望，亦即下列表达式

$$\mathbf{D}(\xi)=\sum_{r=1}^{s}P_r(x_r-\mathbf{M}(\xi))^2$$

称为变量 ξ 的**离差**．离差的平方根

$$\sigma_\xi=\sqrt{\mathbf{D}(\xi)}$$

称为（变量 ξ 与其数学期望 $\mathbf{M}(\xi)$ 的）**均方偏差**．

著名的车比雪夫不等式

$$\mathbf{P}\{|\xi-\mathbf{M}(\xi)|\leqslant t\sigma_\xi\}\geqslant 1-\frac{1}{t^2} \tag{21}$$

就是简单应用离差与均方偏差而得出的．这个不等式告诉我们，ξ 与 $\mathbf{M}(\xi)$ 的偏差显著地大于 σ_ξ 的情形是很难得遇到的．

如果我们作出下列的随机变量的和数

$$\xi=\xi^{(1)}+\xi^{(2)}+\cdots+\xi^{(n)},$$

则对于这些随机变量的数学期望永远有如下的等式：

$$\mathbf{M}(\xi)=\mathbf{M}(\xi^{(1)})+\mathbf{M}(\xi^{(2)})+\cdots+\mathbf{M}(\xi^{(n)}). \tag{22}$$

但对于离差的类似的等式

$$\mathbf{D}(\xi)=\mathbf{D}(\xi^{(1)})+\mathbf{D}(\xi^{(2)})+\cdots+\mathbf{D}(\xi^{(n)}) \tag{23}$$

则仅只是在某些限制之下方才是正确的．为了要使等式(23)成立，例如，具有不同的标码的变量 $\xi^{(i)}$ 与 $\xi^{(j)}$ 之间不是"相关的"便够了，换句话说，对于 $i\neq j$ 有等式[1]

$$\mathbf{M}\{(\xi^{(i)}-\mathbf{M}(\xi^{(i)}))(\xi^{(j)}-\mathbf{M}(\xi^{(j)}))\}=0 \tag{24}$$

便够了．特别，如果变量 $\xi^{(i)}$ 与 $\xi^{(j)}$ 彼此独立[2]，则等式(24)成立，因之，对于彼此独立的加项，等式(23)恒能成立．对于算术平均值

$$\zeta=\frac{1}{n}(\xi^{(1)}+\xi^{(2)}+\cdots+\xi^{(n)}),$$

我们由等式(23)可以推出

$$\mathbf{D}(\zeta)=\frac{1}{n^2}(\mathbf{D}(\xi^{(1)})+\mathbf{D}(\xi^{(2)})+\cdots+\mathbf{D}(\xi^{(n)})). \tag{25}$$

现在我们假定，所有加项的离差不超过某一常数

$$\mathbf{D}(\xi^{(i)})\leqslant C^2.$$

那么根据(25)我们有

$$\mathbf{D}(\zeta)\leqslant\frac{C^2}{n},$$

再利用车比雪夫不等式即知对于任意的 t 我们有

$$\mathbf{P}\left\{|\zeta-\mathbf{M}(\zeta)|\leqslant\frac{tC}{\sqrt{n}}\right\}\geqslant 1-\frac{1}{t^2}. \tag{26}$$

不等式(26)蕴涵了所谓的车比雪夫型的大数定律：如果诸变量 $\xi^{(i)}$

---

1) 我们称表达式

$$R=\frac{\mathbf{M}\{(\xi^{(i)}-\mathbf{M}(\xi^{(i)}))(\xi^{(j)}-\mathbf{M}(\xi^{(j)}))\}}{\sigma\xi_{(i)}\sigma\xi_{(j)}}$$

为变量 $\xi^{(i)}$ 与 $\xi^{(j)}$ 之间的相关系数．若 $\sigma\xi_{(i)}>0$ 且 $\sigma\xi_{(j)}>0$，则条件(24)等价于 $R=0$。

相关系数 R 表达了随机变量之间的依赖程度。我们永远有 $|R|\leqslant 1$，而仅当存在有线性关系

$$\eta=a\xi+b \quad (a\neq 0)$$

的时候才能有 $R=\pm 1$。对于独立的变量，$R=0$。

2) 设两个随机变量 ξ 与 η 分别可取值 $x_1, x_2, \cdots, x_m$ 与 $y_1, y_2, \cdots, y_n$；ξ 与 η 之间的独立性的定义是说，对于任何 i 与 j，事件 $A_i=\{\xi=x_i\}$ 与 $B_j\{\eta=y_j\}$ 在 §2 所给的定义的意义下彼此独立。

彼此无关并且具有有界的离差，则其算术平均值 ζ 当 n 增加时显著地远离它的数学期望 $\mathbf{M}(\zeta)$ 的情形是极其罕见的。

更精确的说法是：如果随机变量序列

$$\xi^{(1)},\ \xi^{(2)},\ \cdots,\ \xi^{(n)},\ \cdots$$

的算术平均值 ζ 于 $n\to\infty$ 时有

$$\mathbf{P}\{|\zeta-\mathbf{M}(\zeta)|\leqslant\varepsilon\}\to 1, \tag{27}$$

（此处 $\varepsilon>0$ 为任意常数），则称此随机变量序列服从大数定律。

为了要从不等式(26)得出极限关系式(27)，只需取

$$t=\varepsilon\frac{\sqrt{n}}{C}.$$

关于怎样尽可能地放宽极限关系式(27)（换句话说，大数定律）的条件的问题，马尔科夫、伯恩斯坦，辛钦以及其他等人作了许多的研究。这些研究具有着原则性的意义。然而还要更加重要的是偏差 $\zeta-\mathbf{M}(\zeta)$ 的概率分布的精确的研究。

概率论方面俄罗斯古典学派的伟大功绩之一是在于确定了这样一个事实，即在非常宽泛的条件之下我们近似地（亦即当 n 无限增加时能以任意大的精确程度断言）有下列公式

$$\mathbf{P}\{t_1\sigma_\zeta<\zeta-\mathbf{M}(\zeta)<t_2\sigma_\zeta\}\sim\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2}e^{-\frac{t^2}{2}}\,dt. \tag{28}$$

车比雪夫在加项独立而且有界的情形下给出了这一公式的差不多完整的证明。马尔科夫补足了车比雪夫的研究中的不充分环节并且放宽了应用公式(28)的条件。李雅普诺夫给出了更为一般的条件。伯恩斯坦研究了公式(28)向具特殊完备性的非独立加项的和数推广的问题。

公式(28)概括了如此大量的特殊问题，因之很久以来它就被称作概率论的中心极限定理。虽然在概率论的最新发展中，它已经被包含在一些更一般的规律之中了，但是即使到了今天它的意义仍然是难以动摇的。

如果加项是独立的并且离差是相同的

$$\mathbf{D}(\xi^{(i)})=\sigma^2,$$

则借助于关系式(25)可将公式(28)表成更方便的形状：

$$\mathbf{P}\left\{\frac{t_1\sigma}{\sqrt{n}}<\zeta-\mathbf{M}(\zeta)<\frac{t_2\sigma}{\sqrt{n}}\right\}\sim\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2}e^{-\frac{t^2}{2}}dt. \qquad (29)$$

现在我们来指明，关系式(29)包含了我们以前考察过的频率 $\frac{\mu}{n}$ 与概率 p 的偏差的问题的解．为此我们引进随机变量 $\xi^{(i)}$，它们的定义如下：

$$\xi^{(i)}=\begin{cases}0, & \text{若第 } i \text{ 次试验得到反面结果,}\\ 1, & \text{若第 } i \text{ 次试验得到正面结果.}\end{cases}$$

很容易验算，此时

$$\mu=\xi^{(1)}+\xi^{(2)}+\cdots+\xi^{(n)},\ \frac{\mu}{n}=\zeta,$$

$$\mathbf{M}(\xi^{(i)})=p,\ \mathbf{D}(\xi^{(i)})=p(1-p),\ \mathbf{M}(\zeta)=p,$$

并且由公式(29)可得

$$\mathbf{P}\left\{t_1\sqrt{\frac{p(1-p)}{n}}<\frac{\mu}{n}-p<t_2\sqrt{\frac{p(1-p)}{n}}\right\}$$
$$\sim\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2}e^{-\frac{t^2}{2}}dt,$$

如果令 $t_1=-t,\ t_2=t$，则上式重新化为公式(20)。

§4. 关于概率论基本概念的补充说明

随机现象的特点在于频率的稳定性，换句话说，特点在于当多次重复实现某一条件组时，从聚集在某一正常水平线——概率 $\mathbf{P}(A/\mathfrak{S})$ ——附近的那些频率中所产生出来的那种稳定的动向；我们在 §1 中讲到这些时，在措词上容忍了两种不确切性．第一种不确切性是在于我们未曾说明，为了要使频率的稳定性一定可以显现出来，各个试验序列中的试验次数 n_r 应大到怎样的程度，以及对于试验次数为 $n_1, n_2, \cdots, n_s$ 的各个试验序列，频率 $\frac{\mu_r}{n_r}$ 彼此之间的能容许的偏差和它们与频率正常水平线 p 之间的能容许的偏差各如何．这种不确切性在一门新科学的概念形成初期是不可避免的．这种不确切性并不更甚于为最简单几何概念所固有的关于点或直线在现实上应如何理解的问题上的某种含糊性．§1 中这一方面的不确切性在 §3 中已经加以确切化

了.

在我们的措词中所隐含着的另一种不确切性是关于那些能使事件 A 的频率的稳定性一定可以呈现的试验序列的构成方法问题.

我们已经看到过，当事件演变无法精确地个别预定的时候，问题才以统计与概率的方法来解决. 所以如果我们希望人为地制造出尽可能纯粹的随机现象，我们就需要特别注意设法，使得没有任何可行的方法能预先分出这样一些情况，在这些情况下现象 A 具有比较某一正常频率更频繁呈现的倾向.

例如，公债抽签可以这样来组织. 假定在公债券总张数为 N 的某一次抽签中，有 M 张公债券被抽中，则个别一张公债券被抽中的概率为 $p=\frac{M}{N}$. 这就意味着，在抽签之前无论用什么方法分出充分多的 n 张公债券，实际上我们总可以确信，这一部分公债券中被抽中的张数 μ 与这一部分公债券的总张数 n 的比值 $\frac{\mu}{n}$ 总是和 p 相接近. 例如愿意买偶数号码的公债券的人并不常比愿意买奇数号码的公债券的人多得到任何好处，完全同样的，相信号码恰能分解为三个质因子的公债券或者号码和前次抽签中被抽中的公债券的号码相邻接的公债券等比较容易被抽中的人也不会多得到任何好处.

在射击的例子里也完全是这样，假如大炮的类型是一定的而且没有任何毛病，炮手是经过良好训练的，炮弹是经过一般产品检查手续，而且用通常方法来取得炮弹，那么与平均落点的偏差小于预先确定的或然偏差 B 的事件约在半数的情形下出现. 在各个射击序列中这个比例皆保持不变，并且如果我们只按偶数次或奇数次（依时间先后计算次数）的射击或按其他类似的办法来计算偏差小于 B 的次数，这个比例也仍然是不变的. 但是，倘若我们选取特别均匀的炮弹（在重量方面等等），那么就完全可能使分散情形缩小，换句话说，对于这一串炮弹，偏差大于标准偏差 B 的将比半数少得多.

因此，只有当说明构成试验序列时可以采取的各种方法的时候，我们才能够说，事件 A 是"概率随机"事件，并且赋予确定的概率

$$p=\mathbf{P}(A/S),$$

而这种说明我们将认为是包含在条件组 S 之内了.

在一定的条件组 S 之下，事件 A 作为概率随机事件并且具有概率 $p=\mathbf{P}(A/S)$ 的这种性质表达了条件组 S 与事件 A 之间的关系的客观特征. 换句话说，不存在绝对随机的事件，事件是随机事件还是必然事件要看是从什么角度来考察，但是在一定的条件之下，事件就可以是完全客观的随机事件，并且它的这一性质不依赖于任何观察者的知识状态. 假定我们设想观察者

能够洞悉炮弹飞行的一切特点与特殊情形，并且因而能够预言每一发炮弹与其平均弹道偏差各为若干，但是即使如此，这位观察者的在场也不会就影响到炮弹散布的服从概率论的规律(当然，假定射击是按照普通方式来进行的，而不是按照我们所设想的这位观察者的指示来进行的)．

关于这一点我们指出，以上我们讨论了如何来构成一个试验序列，使得其中频率能够呈现稳定趋势(意即频率聚集在正常值——概率——的四周)，这种构成也发生在完全不依赖于我们的干涉的实际情形里．例如，正是由于气体分子运动的概率随机特征，甚至在很短的时间区间内，撞击容器壁面的某一部分或撞击放在气体之内的一个物体的表面的分子数目，皆是以很大的精确度正比于这一部分壁面(或物体表面)的面积与时间区间长度．在撞击次数不多的情形下，实际状况和这种正比关系的偏差也服从概率论规律，并且产生布朗运动型的现象，这我们在下文中还要讲到．

现在我们再来说明独立性概念的现实意义．我们回想一下，我们曾经用公式

$$\mathbf{P}(A/B)=\frac{\mathbf{P}(AB)}{\mathbf{P}(B)} \tag{30}$$

来定义在条件 B 下事件 A 的条件概率．我们再回想一下，如果(4)式成立，

$$\mathbf{P}(AB)=\mathbf{P}(A)\mathbf{P}(B),$$

则事件 A 与 B 称为是独立的．从事件 A 与 B 的独立性以及 $\mathbf{P}(B)>0$ 推知，

$$\mathbf{P}(A/B)=\mathbf{P}(A).$$

在概率数学理论中所有关于独立事件的一切定理，其应用对象总是满足条件(4)(或满足其在多个事件彼此独立的情形下的推广)的任何事件．然而独立性的这个定义如果不是由于(因果意义上的)现实的独立现象的特性而产生的，那么这些定理也就没有什么意思了．

例如，大家知道，新生婴儿为男婴的概率有着相当稳定的值 $\mathbf{P}(A)=\frac{22}{43}$．如果以 B 表示这样一个条件，即诞生之日恰逢木星与火星相合，那么假定行星的交互位置并不决定人的个别命运，条件概率 $\mathbf{P}(A/B)$ 即应有同样的值 $\mathbf{P}(A/B)=\frac{22}{43}$，这就是说，若是在这种特殊的星相条件之下来计算诞生男婴的频率，则仍会得到频率 $\frac{22}{43}$．尽管可能谁也没有充分大规模地这样计算过，但是我们没有理由怀疑这种计算结果．

我们引进了这个就内容说多少已经陈旧了的例子为的是表明，人类知识的发展不仅在于确定现象之间的真实联系，而且还在于批驳那些虚构的联系，也就是说在适当的情形下建立关于某两种范畴的现象之间的独立性的命

题．揭破星相家妄图把本无联系的两种范畴的现象强行联系起来的荒谬性，就是这种古典的事例之一．

自然，这种独立性也不应该绝对化．例如，由于万有引力定律，木星卫星的运转毫无疑问就会对于(例如)炮弹飞行产生一些影响．但是显而易见，实际上我们可以不考虑这些影响．从哲学观点看来，不说独立性而说某种依赖性在一定具体情况下是非本质的也许更恰当一些．不过无论如何，在现在所阐释的具体的与相对的意义下的事件独立性，丝毫也不会和一切现象普遍联系的原则相抵触，而仅只是这一原则的必要的补充．

在事件开始原是独立的，随后由于现象的演变而转相联系的情况下，利用由一些事件的独立性假定出发而推出的公式来计算概率是有现实价值的．例如，当计算宇宙放射微粒与它们所穿过的介质中的微粒相碰撞的概率时，可以假定在高速运动着的宇宙放射微粒出现在介质微粒的近傍之前，介质微粒的运动不依赖于这些宇宙放射微粒的动态．当计算敌人的子弹打中旋转着的螺旋桨的桨叶的概率时，可以假定对中心轴而言桨叶的相对位置不依赖于子弹的轨道(当然，这个假定对于随着螺旋桨的转动而瞄准螺旋桨放射的那些特殊的子弹来说是错误的)；象这样的例子不胜枚举．

我们甚至可以说，在概率规律性呈现得十分明显的一切地方，我们所遇到的总是或是彼此完全独立或是在某种意义下微弱地相联系的大量因素的协同作用．

这绝对不是说，我们可以到处不严格地引入这种或那种的独立性的假定．反之，这种情况迫使我们：第一，必须特别仔细地来研究独立性假设检验的准则；第二，必须特别仔细地来考察边界情形，在这些边界情形中各因素之间的依赖性必然需要加以考虑，但是同时这些依赖性还不够强，概率规律性在变化了的复杂化了的形式下还可以呈现．在前文中我们曾指出，概率论方面的古典俄罗斯学派已经广泛地研究了这第二个方面．

在结束我们关于独立性问题的考察时，我们指出，无论是由公式(4)所给出的两个事件间的独立性的定义，还是多个随机变量间独立性的形式定义，其含义均较(按照属于无因果联系的不同现象范畴的这种意义来了解的)现实的独立性概念广泛得多．

例如，假定一个点 ξ 按照下述的规律落在区间 $[0, 1]$ 之上，即对于任何

$$0 \leqslant a \leqslant b \leqslant 1,$$

点 ξ 落在区间 $[a, b]$ 上的概率等于这个区间的长度 $b-a$．不难证明，此时若把点 ξ 的坐标值按十进位分数展开：

$$\xi = \frac{a_1}{10} + \frac{a_2}{100} + \frac{a_3}{1000} + \cdots,$$

则各个 a_k 是彼此独立的，虽然按 a_k 的产生而言他们是彼此相联系的[1]。(由此可以推出许多有理论价值的推论，其中也有一部分具有实际的意义.)

独立性形式定义的这种柔韧性不应看成是它的缺点. 相反地，它只是把在关于独立性的这种或那种假定之下建立的定理的应用范围扩大了. 它使得无论独立性是从实际考虑出发而加以假定的，还是从所考察的事件与随机变量的概率分布的预定假设出发而以计算方法证实的，这些定理都同样地可以应用.

一般言之，研究概率论的数学工具的形式结构，可以推出一些有意思的结果. 原来这一工具在近世数学的基本研究对象的逐渐形成的分类中占有十分明确的而且非常质朴的地位.

我们已经讲过了两个事件的兼有 AB 的概念以及两个事件的联合 $A\cup B$ 的概念. 我们再重复一下，如果事件 A 与 B 的兼有是不可能的，亦即如果 $AB=N$ (此处 N 表示不可能事件)，则我们称它们是不相容的.

初等概率论的基本公理之一是要求(参看 § 2) 当条件 $AB=N$ 成立时恒有等式

$$\mathbf{P}(A\cup B)=\mathbf{P}(A)+\mathbf{P}(B).$$

概率论基本概念——随机事件及其概率——的性质完全类似于平面图形及其面积的性质. 我们完全可以把 AB 理解为两个图形的交叉(公共部分)，把 $A\cup B$ 理解为两个图形的合并，把 N 理解为习惯引用的"空"图形，而把 $\mathbf{P}(A)$ 理解为图形 A 的面积，并且在上面所指出的范围内这种类似性是充分的.

三维图形的体积也具有同样的性质.

目前这种类型的结构的最一般的理论(包括体积与面积理论作为特殊情形)是一般**测度论**，关于测度论的若干知识可以在第十五章(第三卷)实变函数论中找到.

只是应当指出，概率论和一般测度论来比较，或者专门和面积与体积理论来比较，自有其独具的若干特征：概率永远不会大于 1. 必然事件 U 具有这个最大的概率：

$$\mathbf{P}(U)=1.$$

类似性并非只是表面上的. 原来从形式观点看来，全部概率数学理论可

---

1) 把点 ξ 的坐标值按照任何 n 进位分数而展开，

$$\xi=\frac{a_1}{n}+\frac{a_2}{n^2}+\frac{a_3}{n^3}+\cdots,$$

这一结果都能成立.

以构造成以“整个空间 U 的测度为 1”这一假定特殊化了的测度论[1]。

这样看待问题不仅使得概率数学理论的形式结构显得十分清晰，而且还使得概率论本身及形式结构与之相近的其他数学理论都获得了非常实际的进展。在概率论中有成效地使用了实变函数度量理论中探讨出来的精微的方法。同时在邻近数学领域的问题中，概率方法不仅可以“触类旁通”式地加以袭用，而且可以形式严整地原封不动地移用于新的领域。随便什么地方，只要那里概率论公理能够成立，那里就可以引用这些公理的推论，即便是那里和现实的随机性没有任何共同之点也可以不管。

由于有了公理化的概率论，就使得我们摆脱了试图寻找一种既具有自然科学的直观确凿性又便于建立形式严整的数学理论的方法来定义概率的诱惑。这样的定义大体上有如我们在几何学中把点这样来定义，即把点作为一个实在的物体经过四面八方无数次切削（而且每次切削均使得直径缩小（例如）一半）最后所剩余下来的东西。

把概率定义为试验次数无限增大时频率的极限，就是属于这一类的定义法。如果我们假定试验具有概率特征，换句话说，假定频率有聚集在一个常数值周围的倾向，那么这种假定本身仅只是在某些条件之下方才能够成立，但是事实上我们无法无限长久地无限精确地保持这些条件。因此精确的极限过程

$$\frac{\mu}{n}\to p$$

不可能有现实意义。如果要把频率稳定的原理归结成为这样的极限过程，那么就需要确切定出寻找无穷试验序列的可行的方法，而这种无穷试验序列也只能是数学上的虚构。如果最后我们所得到的理论非常的特别，以致用其他任何方法都不能给出这个理论的严格论证，那么整个这一系列的概念可能还需要加以严肃的审查。但是，以上我们已经指出，在测度论的现状之下，为要论证概率数学理论只需补加以下条件

$$\mathbf{P}(U)=1$$

即可。

一般而言，当实际分析概率概念时，不一定就会导致它的形式定义。显然，从纯形式观点看来，关于概率最多我们只能这样说：概率 $\mathbf{P}(A/S)$ 是一个数，当实现条件组 S 并按条件组 S 所规定的方法构造试验序列时，频率有聚集在这个数的周围的倾向，如果我们在不破坏条件的一致性的合理范围内增

1) 尽管如此，依照所解决的问题的实质看来，概率论仍是一门独立的数学科目；某些结果（如 §3 中所讲的那些）对于概率论说来是基本的，但从纯粹测度论观点看来，却似乎是人为制造出来的，似乎是用不着的。

大这些序列的数目，并且使我们的所有的序列都具有(对于给定的具体情况而言的)充分的可靠性与精确性，那么这种倾向就能够呈现得任意清楚与准确.

实际上，重要的问题倒不在于这个定义在形式上应如何来确定，而是在于使这种概率随机性得以显现的条件应如何尽可能广泛地加以阐明. 应该清楚地了解到，事实上直接以统计核验方法来论证某种现象的概率特征的假定是非常稀见的事. 只有当在某个新的领域中初次施用概率方法的时候，才时常是首先用纯经验的方法来查明频率的不变性. 根据我们在§3中所讲的，为了要查明精确度为ε的频率不变性，试验序列就必须包含大约$n=\frac{1}{\varepsilon^2}$次试验. 例如，为了要肯定，在某一具体问题中考虑精确度为0.0001的概率是有意义的，那么我们就必须要有许多个包含试验多达100,000,000次左右的试验序列.

引进概率随机性的假定的更常见得多的方法是考虑对称性或是考虑和问题有关的各种现象间的实际独立性等等，然后再用间接办法来检验这个假定. 例如，由于在有限体积气体中分子的数目常常是以10^{20}来计量或者更多，因之在气体动力学的许多概率结果中相应的$\sqrt{n}$时常是很大的，而且实际上这些结果中有许多是具有极大精确性的. 例如，在悬挂在静止空气中的一张薄板上，尽管对于一定的实验设计而言其两面所受压力只要相差0.001%就可以觉察出来，但是甚至在目不能见的一小块面积上其两面所受压力仍然是精确地相等.

§5. 因果过程与随机过程

所有现象因果制约的原理在利用微分方程研究实际过程的方法中得到了最简单的数学表示；这种方法已在第五章§1的一系列的例中给出示范.

假定我们所考察的体系在时刻t的状态可由n个参数

$$x_1,\ x_2,\ \cdots,\ x_n$$

来确定。这些参数的变化速度，众所周知，可以用它们对于时间的导数

$$\dot{x}_k=\frac{dx_k}{dt}$$

来表示.

如果我们设这些速度是各个参数的函数，则我们就得到了一个微分方程组：

$$
\begin{aligned}
\dot{x}_1 &= f_1(x_1,\ x_2,\ \cdots, x_n),\\
\dot{x}_2 &= f_2(x_1,\ x_2,\ \cdots,\ x_n),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\dot{x}_n &= f_n(x_1,\ x_2,\ \cdots,\ x_n).
\end{aligned}
$$

在数学化的自然科学发生的初期所发现的自然法则的绝大部分，从关于落体的伽里略定律开始，都能以这种形式来表达. 伽里略未能赋予他的这一发现以上述的标准形式，因为彼时相应的数学概念尚未发现. 这一工作系由牛顿所完成.

在力学以及物理学的许多其他范畴中，常出现二阶的微分方程，从原则上说，这并未引进任何新的东西，因为倘若我们对于速度 $\dot{x}_k$ 采用新的符号

$$v_k = \dot{x}_k,$$

则变量 x_k 的二阶导数有如下的表达式：

$$\frac{d^2 x_k}{dt^2} = \dot{v}_k,$$

并且 n 个变量 $x_1,\ x_2,\ \cdots,\ x_n$ 的二阶方程可以转化为 $2n$ 个变量 $x_1,\ x_2,\ \cdots,\ x_n,\ v_1,\ v_2,\ \cdots,\ v_n$ 的一阶方程.

作为一个例子，我们试来考察一件重物在地面大气中下落的问题. 假定在我们的问题中所考察的距地面的高度是不大的，那么我们可以认为介质阻力只依赖于落体的速度而不依赖于高度. 我们所考察的体系的状态可以由物体与地面的距离 z 及其速度 v 这两个参数来决定. 这两个参数在时间过程中的变化可以由如下的两个微分方程

$$
\begin{aligned}
\dot{z} &= -v\\
\dot{v} &= g - f(v)
\end{aligned}
\qquad (31)
$$

来决定，这里 g 表示重力加速度，$f(v)$ 表示对于我们所考察的物体的某种“阻力法则”.

如果速度不大，而且物体充分重，例如中等大小的石头自若干公尺的高度落下时就是这种情形，此时大气阻力可以不计，而方程(31)可化为

$$
\begin{aligned}
z &= -v, \\
v &= g.
\end{aligned}
\tag{32}
$$

如果假定变量 z 与 v 在初始时刻 t_0 有值 z_0 与 v_0，则方程(32)不难解出，从而得到描述整个下落过程的下列公式：

$$
z = z_0 - v(t-t_0) - g\left(\frac{t-t_0}{2}\right)^2.
$$

例如，假定 $t_0=0$, $v_0=0$，我们便得到了伽里略所发现的下列公式：

$$
z = z_0 - \frac{gt^2}{2}.
$$

在一般情形下，方程(31)的积分问题稍复杂一些，但是原则性的结果（在对于函数 $f(v)$ 的极其一般的限制之下）则仍然是同样的：根据变量 z 与 v 在初始时刻 t_0 的值 z_0 与 v_0，可以单值地算出 z 与 v 在 t_0 以后物体落到地面以前任何时刻 t 的值．如果我们假定 z 取负值时仍能继续下落，则“物体落到地面以前”这一限制也可以在想象上予以取消．对于用这种办法概型化了的问题，我们可以作出如下的论断：如果函数 $f(v)$ 当 v 增加时单调递增且当 $v\to\infty$ 时趋于无穷，则当无限地继续下落时，换句话说当变量 t 无限增加时，速度 v 必趋于固定的极限值 c，这个极限值 c 是下列方程

$$
g = f(c)
$$

的根．

就直观而论，上述问题的这个数学分析的结果是很容易了解的：下落速度达到重力加速度不复能抵消大气阻力时就会停止增加．当携带着张开的降落伞跳下时，很快就可以达到[1]稳定速度 v（约每秒五公尺）．当携带着未张开的降落伞持续下落时，大气阻力减少，因之稳定速度增高．并且只有当跳伞人降落了很大距离

1) 确切些说是 v 变得和 c 实际上充分近．

之后才能达到稳定速度.

如果是一个轻的物体下落，例如在大气中抛下一根羽毛或绒毛，则能够看出的加速度运动的初始阶段是很短的，而且有时完全是看不出来的．下落的稳定速度很快就能达到，在一定近似程度上，我们可以在整个过程中取 $v=c$. 于是此时就只剩下一个微分方程了：

$$z=-c,$$

这个微分方程可以很简单地积分出来，

$$z=z_0-c(t-t_0).$$

在完全静止的空气中抛下一根绒毛的情形就是这样.

以上我们着重指出的因果概念在动力系统近代理论中可以讲述成非常一般的形式，苏联数学家波果留博夫，斯捷巴诺夫以及其他许多人在这一理论中有着许多重要的贡献．从这个一般性的理论看来，甚至系统状态已经不是由有限个参数来决定而是要由一个或几个函数来决定的那些实际过程(例如，连续介质力学中的情形就正是这样)的数学概型，也仅只是一些特殊情形．在这些情形中，系统状态在“无穷小的”时间内的变化的基本规律已经不能由常微分方程来描述，而是要用偏微分方程或者其他的方法．但是实际过程的所有的因果性的数学概型具有如下的两个共同点：第一，我们所考察的系统的状态可以看成是由于给定了某一数学对象 ω (例如 ω 可以是 n 个实数所构成的组，也可以是一个或几个函数等等)的值而引起的各种表象的总体；第二，ω 在时刻 $t>t_0$ 的值由其在初始时刻 t_0 的值 ω_0 唯一决定：

$$\omega=F(t_0,\ \omega_0,\ t).$$

我们在前面已经看到过了，对于能用微分方程来描述的那些过程，要找出函数 F 需要在初始条件“$t=t_0$ 时 $\omega=\omega_0$”之下来积分这些微分方程.

机械唯物论的一些代表人物认为，上述的概型乃是实际现象的宿定性(因果性物理原理)的精确而且完全的表现．按照拉普拉斯的意见，宇宙在给定时刻的状态可由适合无穷多个微分方程的

无穷多个参数来决定．假如有某一个"无所不知的大天才"能够写出所有这些方程，并且能够把它们积分出来，那么依拉普拉斯看来，他就能够完全准确地预见宇宙在无穷时间过程中的全部演化．

然而实际上，量的方面的数学的无穷性比起现实世界的质的方面的无涯无尽性来是极为粗浅的．无论是引进无穷多个参数，还是利用空间的点函数来描述连续介质状态，都不是实际现象的无限复杂性的等价的反映．

正如第五章§3中所着重指出的，实际现象的研究并不总是朝着增加问题中所引用的参数的数目这一方向发展；一般说来，在为了考虑某一现象而采取的数学概型中，把决定个别"系统状态"的特征标志ω复杂化决不能说是永远都合宜的．恰恰相反，研究者的技巧倒正是在于寻找一个非常简单的位相空间Ω（即ω的值的集合，换句话说，即系统的各种可能状态的集合）[1]，使得当我们把实际过程换成点ω在这个空间中的因果式的变迁过程时，仍能抓住实际过程的各个主要方面．

但是，从实际过程中抽出了它的主要特征之后，就会遗留下一些残余的东西，这些残余我们只好算作是随机的．在过程的演进中，这些未纳入考虑范围的随机因素总会产生一些影响．用数学研究一个现象，而最后当把理论结果和实际观察结果相比较时，竟然不能觉察这些未纳入考虑范围的随机因素的影响，这是非常罕见的情形．关于引力作用下行星运动的理论就正是这种情形或者是近似于这种情形：行星间的距离比较起行星本身的大小来是如此之大，使得我们几乎永远可以把它们理想化为质点；在行星所经历的空间中充满了如此稀薄的物质，使得阻力小到几乎没有；行星的质量又是如此之大，使得在它们的运动中光压几乎不起作用；由于有了这样一些难得的情况，所以，n个质点的系统（其"状态"可由$6n$个参数来描述）[2]在单纯引力作用之下的运动问题的数学解，

1）在以上所引的物体下落的例子中，位相空间Ω是数对(s, v)的集合，换句话说就是平面．关于位相空间的一般考察可参看（第三卷）第十七章及第十八章§3．

2）换句话说，可以用每个点的三个坐标与三个分速度来描述．

能够以如此惊人的准确程度与行星运动的观察结果相符合。

如果在运动方程中计及大气阻力，则炮弹飞行的情形和行星运动的情形有一定程度的相近。这也是数学方法能够比较容易地与迅速地取得显著成果的那些古典问题之一。但是在这个问题中，各种随机因素的干扰作用已经要大得多了，并且炮弹的散布，换句话说，实际弹道和对应于该次射击所给定的正常初始条件以及射击时大气平均状态的理论弹道之间的偏差，常达数十公尺，而在远距离射击时常达数百公尺。这种偏差之所以产生，一部分是由于初始射向的随机偏差以及初始速度与正常速度之间的随机偏差，一部分是由于炮弹质量及阻力系数的随机偏差，还有一部分是由于实际地面大气层中的不均匀性与风的突然发生以及大气层中其他具有决定性作用的极为复杂，变化莫测的随机因素。

炮弹的散布可以用概率论的方法来详细研究，并且这种研究的结果对于实际射击是极其重要的。

但是随机现象的研究到底是怎么一回事呢？看起来，当现象依照某一方法而进行概型化的时候，如果随机的"残余"大到不可忽略的程度，那么似乎唯一可能的办法就是引入一些新的参数来更细致地描述我们的现象，并且在这个复杂化了的概型中更详尽地来进行研究，犹如这就是因果现象的数学概型一样。

这种办法在多数场合下是行不通的。例如，在研究物体于空气中下落时，若是要考虑到不均匀的突发的（或按通常的说法，湍性的）气流的构成，那么以前所用的两个参数 z 与 v 就不济事了，为了要完全描述出这种构成，需要引用的数学工具简直是浩如烟海，不可胜计。

但是事实上，只有当我们无论如何必须十分详细地对于每一个别情形逐一地考察过程的演进中残余"随机"因素的影响时，这种复杂的方法才是不可避免的。幸而，我们的实际需要常常完全不是这样，而仅只是需要估计在一个长时间内或大量重复我们所考察的过程时随机因素作用的总影响。

作为一个例，我们试来考察河流中或是某一人工水利工程建

筑中水的流动所造成的沙的迁徙．这种迁徙通常是这样发生的：大多数砂粒都安安静静地躺在水底下，而仅只是间或有一些靠近水底的特别强有力的旋涡卷走一些个别的砂粒，而且一下子就把它们带过一个相当长的距离，一直到某一个新的地方突然停下为止．每一个这样的砂粒的纯理论运动，可以分别根据水力学中的种种定律来研究．但是为此我们就必须详尽地了解底部及水流的初始状态，然后再逐步地来研究这种运动：先指出在任何一个被卷走的砂粒上，压力什么时候方才大到足以使其运动，再详细研究运动的砂粒在其停止之前的全部迁徙情况．这样来提问题，对于实际科学研究显然十分荒谬．虽然如此，但是关于水流所造成的底部沉积物的迁徙的平均规律性(或按照普通的说法，统计规律性)却完全是可以研究的．

由于大量随机因素协同作用的结果而产生非常显著的统计规律性的实例可以很容易地举出许多来．最有发展希望的与最为人所熟知的这种实例之一就是气体运动学，它告诉了我们气体作为一个整体对于器壁的压力以及一种气体往另一种气体中的扩散等所具有的那些精确的规律性是如何从分子的许多随机碰撞的协同作用而产生的．

§6. 马尔科夫型的随机过程

在§5中我们曾以方程

$$\omega=F(t_0,\ \omega_0,\ t)$$

来描述因果概型，把这种概型直接推广成为概率概型的功绩系属于马尔科夫．固然，在马尔科夫所研究的情形中，所考察的系统的位相空间仅只是包含有限个状态：$\Omega=\{\omega_1,\ \omega_2,\ \cdots,\ \omega_n\}$，并且仅只是把时间 t 分成一些离散的步骤来研究系统状态的变化．但是他善于从这个极端概型化的模型中找出许多基本的规律．

在因果概型中我们有一个函数 F，借助于这个函数 F 可以根据时刻 t_0 的状态 ω_0 唯一地确定时刻 $t>t_0$ 的状态 ω，而马尔科夫

则与此相应地引入在"于时刻 t_0 处于状态 ω_i" 的条件下 "于时刻 t 出现状态 ω_j"的概率

$$\mathbf{P}(t_0,\ \omega_i;\ t,\ \omega_j).$$

对于三个不同的时刻

$$t_0<t_1<t_2,$$

这种概率之间的关系，马尔科夫用下式表达出来：

$$\mathbf{P}(t_0,\ \omega_i;\ t_2,\ \omega_j)=\sum_{k=1}^{n}\mathbf{P}(t_0,\ \omega_i;\ t_1,\ \omega_k)\mathbf{P}(t_1,\ \omega_k;\ t_2,\ \omega_j). \tag{33}$$

此式可以称为马尔科夫过程基本方程.

当位相空间是连续流形时，最标准的情形就是存在有一个从状态 ω_0 出发经过时间区间 (t_0, t) 变到状态 ω 的概率密度 $p(t_0, \omega_0; t, \omega)$. 在这种情形下，从状态 ω_0 出发经过由时刻 t_0 到时刻 t 这一段时间变到位相空间 Ω 的子区域 G 中的任何状态 ω 的概率可以写成下列形状：

$$\mathbf{P}(t_0,\ \omega_0;\ t,\ G)=\int_G p(t_0,\ \omega_0;\ t,\ \omega)d\omega, \tag{34}$$

此处 $d\omega$ 是位相空间 Ω 中的容积元素[1]. 对应于概率密度 $p(t_0, \omega_0; t, \omega)$，基本方程(33)采取下列形状：

$$p(t_0,\ \omega_0;\ t_2,\ \omega_2)=\int_\Omega p(t_0,\ \omega_0;\ t_1,\ \omega)p(t_1,\ \omega;\ t_2,\ \omega_2)d\omega. \tag{35}$$

要来解方程(35)那是非常困难的，但是在某些限制之下可以从方程(35)导出一些比较易于研究的偏微分方程来. 物理学家佛开尔与普兰克从不严格的物理考虑得到了一些这样的方程. 而所谓的随机微分方程的理论的全貌则系由若干苏联学者(伯恩斯坦、柯尔莫果洛夫、彼得罗夫斯基、辛钦等)所建立.

我们不想在这里引入这种方程.

例如，假定有一个很小的物体，由于它的体积太小，空气分子

1) 严格说来，等式(34)乃是概率密度的定义. 数值 $pd\omega$ 等于(如果忽略高级无穷小)从状态 ω_0 出发经过由 t_0 到 t 这一段时间变到容积元素 $d\omega$ 中的概率.

在它的对立面的推动力不能完全均衡，从而产生了“布朗运动”，并且它的平均下落速度 c 比起它的布朗运动速度来要小得多，那么对于这种物体在静止的大气中的运动问题就可以利用随机微分方程的方法很容易地得到解决。

设 c 是平均下落速度，D 是所谓扩散系数。如果我们假定，质点到达地面($z=0$)时并不停滞而是产生“反射”，换句话说，在布朗力的作用之下重新又进入大气游动，同时再假定质点于时刻 t_0 处于高度 z_0，那么质点于时刻 t 处于高度 z 的概率密度 $p(t_0, z_0; t, z)$ 可由下列公式表出：

$$p(t_0, z_0; t, z)=\frac{1}{2\sqrt{\pi D(t-t_0)}}\left[e^{-\frac{(z-z_0)^2}{4D(t-t_0)}}+e^{-\frac{(z+z_0)^2}{4D(t-t_0)}}\right]\times e^{-\frac{c(z-z_0)}{2D}-\frac{c^2(t-t_0)}{4D}}+\frac{c}{D\sqrt{5}}e^{-\frac{cz}{D}}\int_{\frac{z+z_0-c(t-t_0)}{2\sqrt{D(t-t_0)}}}^{\infty}e^{-z^2}dz.$$

在图 4 中表出了曲线 $p(t_0, z_0; t, z)$ 对于顺序的各个时刻 t 的变化情形。

我们看到，质点的平均高度是递减的，而其位置则越来越不确定(越来越“随机”)。最有意思的是当 $t\to\infty$ 时对于任何 t_0 与 z_0

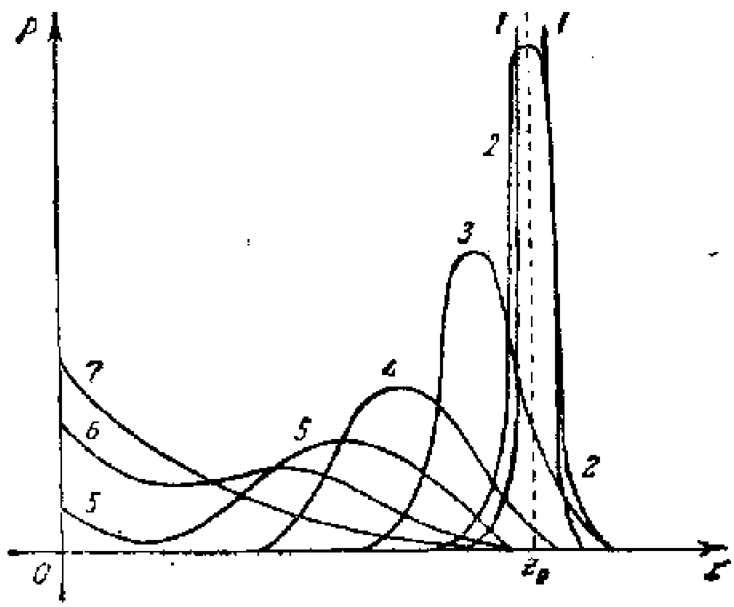


图 4

恒有

$$p(t_0,\ z_0;t,\ z)\to\frac{c}{D}e^{-\frac{cz}{D}},\tag{36}$$

换句话说，对于质点的高度存在有极限分布，并且质点所在的高度的数学期望当 $t\to\infty$ 时趋于如下的正的极限：

$$z^*=\frac{c}{D}\int_0^\infty ze^{-\frac{cz}{D}}dz=\frac{D}{c}.\tag{37}$$

这样一来，地面以上的一个质点尽管受了重力作用随时都有下落的倾向，但是当这个过程(即在空气中的游动)无限延续的时候，它却平均处于一个一定的正的高度．倘若我们取初始的 z_0 小于 z^*，那么可以断言，经过了充分长的一段时间之后，质点的平均位置将高于它的初始位置，这种情形可由图 5 (其中 $z_0=0$)看出．

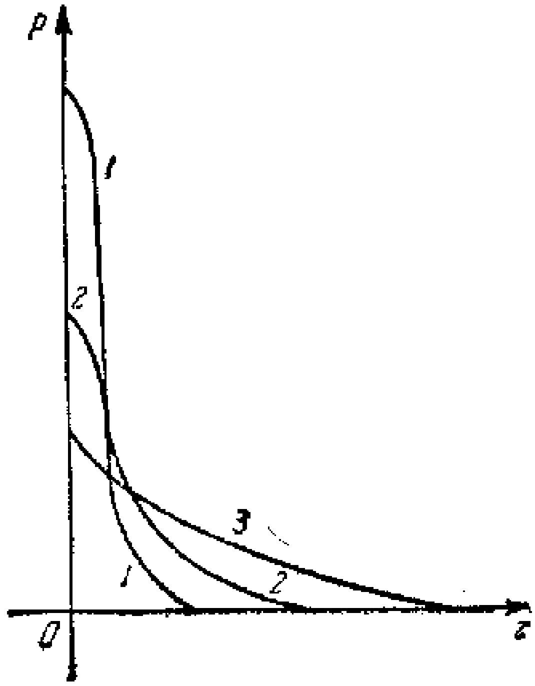


图 5

对于个别的质点，上述的平均值 z^* 仅只是数学期望，但是根据大数定律，对于大量的质点它们就是实际可以实现的了：这种质点在高度方面的分布密度服从前述的规律，并且特别言之，经过充分长一段时间之后，分布密度将按照公式(36)而趋于稳定．

以上所说种种仅只是直接适用于掺杂在空气中的浓度不大的气体，烟等等，因为数值 c 与 D 系假定由预先给定的大气状态所确定的．然而把理论修改得更细致一些之后，就可以适用于构成大气的各种气体之间的彼此扩散，以及由于这种扩散的结果而产生的各种气体就高度而言的密度分布．

比值 $\frac{c}{D}$ 随着质点体积的增大而递增，并且由于这个缘故，质

点混合过程即由具有扩散的性质而变为具有按§5所讲的那种规律而规律性下落的性质. 应用我们的理论，可以检明纯粹扩散运动与这种规律性下落之间各种中间情形。

悬浮在大气中的质点在湍性混合作用之下的运动问题更要复杂一些，但是原则上亦可运用类似的概率方法。

文　献

通 俗 读 物

Б. В. Гнеденко и А. Я. Хинчин, Элементарное введение в теорию веро-ятностей, Изл. 3-е, Гостехиздат, 1952.

现代的大学教程

С. Н. Бернштейн, Теория вероятностей, Изд. 4-е, Гостехиздат, 1946.
Б. В. Гнеденко, Курс теории вероятностей, Изд. 2-е. Гостехиздат, 1954.
格涅坚科，概率论教程，高等教育出版社.

专　著

С. Чандрасекар, Перенос лучистой энергии, М., 1953.
А. Эйнштейн и М. Смолуховский, Броуновское движение, Сборник статей, 1936.

梁文骐 译
王寿仁 校

第十二章　函数逼近法

§1. 绪　论

在实际生活中，我们经常要把一个数用另外的数来逼近．只需说明，当测定这种或那种具体的量(长度、面积、温度等等)时，我们所得到的仅仅是近似地表示这些量的数．实际我们只用到有理数，亦即形如$\frac{p}{q}$的数，在这里p与$q(q\neq 0)$都是整数．但是除了有理数以外还有无理数，并且如果我们在测定具体的量时得不到无理数，而由理论上的推演却往往得到这样的数．例如我们知道半径为$r=\frac{1}{2}$的圆的周长等于无理数π，勾及股长都是1的直角三角形的弦长等于$\sqrt{2}$．在计算无理数的运算时，首先以必需的准确度把它们用相应的有理数来逼近，通常用十进位有限小数来逼近．

对于函数也有同样的情况．数量性的自然规律不是用函数准确地反映在数学上，而是以这种或那种准确度反映在数学上．不但如此，在许多情形下，为了实际计算起见，对于完全用数学法则所给出的已知函数，需要以一定的准确度用其它的函数来逼近．

然而问题不仅在于计算．一函数用其它函数近似表示的问题有很大的理论上的意义．这可以用几句话解释如下．在数学分析发展的过程中，已经发现过并且研究过很重要的逼近函数类，即是那些函数，在熟知的条件下逼近其它多少任意的函数的自然工具．首先代数及三角多项式是这种函数类，它们的各种推广也是这样．我们知道根据要逼近的已给函数的性质，在一定的条件下可以判断它与逼近它的函数的偏差的性质．反过来说，知道了偏差的性质，例如知道了已给函数与逼近它的函数所组成的一定级数的偏

差的大小，就可以知道已给函数的性质．在这条道路上，以函数用这种或那种逼近函数类所作的近似表示为基础，已经建立了一种函数论．在数论中也有类似的理论．在这种理论中，无理数是用逼近它的有理数来进行研究的．

在第二章(第一卷)中，读者已经认识了一种很重要的逼近方法，即泰勒公式．如果函数满足一定的条件，利用这种公式，可以把这函数用其它形如 $P(x)=a_0+a_1x+\cdots+a_nx^n$ 的函数、亦即所谓代数多项式来逼近．在这里 a_k 是与 x 无关的常数．

代数多项式是构造很简单的函数；要按已给系数 a_k 与已给值 x 来计算它，只需应用三种算术运算(加法、减法及乘法)．计算简单在实用上是极其重要的．代数多项式之所以用作逼近函数的最普遍的工具，这就是原因之一(以后再讲另一重要的原因)．只需说明，特别在近来我们必须大规模地在计算机上作技术计算．现代完善的计算机能够很迅速地、不息地进行计算．然而机器只能作比较简单的运算．我们可以使它作非常大量的算术运算，可是譬如说不能使它最后实现自变量趋近于极限的无穷过程．例如机器不能准确地计算 $\lg x$．不过我们可以以必需的准确度用多项式 $P(x)$ 逼近 $\lg x$，然后用机器计算多项式．

除了泰勒公式以外，还有在实用上有重大意义的、用代数多项式逼近函数的其它方法．在这方面首先特别是在积分的近似计算中以及在微分方程的近似积分中所广泛使用的各种插值法．在平均平方的意义下的逼近法得到了更大的推广，这种方法不仅广泛被应用到代数多项式上．伟大的俄国数学家车比雪夫所提出的最好一致或车比雪夫逼近法在一定的实用领域内有重大的意义，如以后我们所要看到，这种方法是由关于机械的构造问题所产生的．

我们的问题是提出这些方法，并且尽可能指出一些条件，在这些条件下，一种方法应当比另外的方法好．没有一种方法是绝对较好的．每种方法在一定的条件下可以比其它方法好．例如如果讲到解决物理问题，那么对所用到的函数究竟应当用这种或那种逼近法，这往往是由问题的本质、或如所谓是由物理上的理由所决

定的.

我们也将看到,在一定的情况下,可以应用一种方法;而不能应用其它方法.

所讲到的每种方法是在适当的时期出现的,有它本身的理论与历史.牛顿已经知道了插值公式,并且用所谓差分比给出了这公式在实际计算中很适用的表示式.在平均平方意义下的逼近法至少有150年的历史.可是在长久一段时期中,这些方法在理论上没有任何联系.它们仅仅是逼近函数的各别的实用方法,而且这些方法应用的范围也是不清楚的.

现在的函数逼近论是在车比雪夫的著作以后出现的.这位数学家引进了在科学上重要的最好逼近、特别是一致最好逼近的概念,系统地把它应用到其它问题中,并且研究了它的理论基础.最好逼近是近代函数论中所运用的基本概念.在车比雪夫以后,他的思想在他的学生佐洛塔辽夫、柯尔金以及马尔科夫兄弟的著作中得到了进一步的发展.

在车比雪夫的时期,除了引进了基本的概念以外,在对任意个别函数求最好逼近的方法上函数逼近论得到了发展,我们在现代要广泛用到这些方法;此外,从与必须逼近函数有关的需要这种观点出发,当时对逼近类的性质,首先是代数及三角多项式的性质,奠定了研究的基础.上世纪末德国数学家魏尔斯特拉斯在数学上的重要发现,对函数逼近论的进一步的发展有影响.他完全严格地证明了:在原则上可以用代数多项式以任意给定的准确度逼近任意的连续函数.代数多项式之所以是函数逼近的普遍工具,这就是第二个原因.要把代数多项式用作这种工具,仅仅知道它们的结构简单还不够;还必须在原则上可以用它们以预先给定的任意小误差逼近连续函数.魏尔斯特拉斯证明了这是可能的.

车比雪夫的最好逼近的深刻思想以及魏尔斯特拉斯定理是函数逼近论的现代新方向的发展(从本世纪初开始)基础.与这相关,我们提出伯恩斯坦、波雷耳、杰克逊、勒贝格以及瓦雷-普森的名字.这种方向可以简单说明如下.在本世纪初与在车比雪夫的时

期(本世纪初以前)一样,通常只提出个别函数的逼近问题;而现代用多项式或其它逼近函数的逼近问题有它的特点:被逼近的函数不是个别给出的函数,而是或多或少广泛的函数类(解析函数类、可微分函数类等等)中任一函数.

俄国的以及现在苏联的逼近论学派在历史上起着主导的作用.苏联的数学家伯恩斯坦、柯尔莫果洛夫、拉夫伦捷夫以及他们的学生们对逼近论作了重大的贡献.实质上,这种理论在现代已经发展成为函数论的一个独立分支.

除了代数多项式以外,三角多项式是逼近函数的很重要的工具.形如

$$u_n(x)=\alpha_0+\alpha_1\cos x+\beta_1\sin x+\alpha_2\cos 2x+\beta_2\sin 2x+\cdots$$
$$+\alpha_n\cos nx+\beta_n\sin nx$$

或简单写成

$$u_n(x)=\alpha_0+\sum_{k=1}^{n}(\alpha_k\cos kx+\beta_k\sin kx)$$

的函数称为 n 阶三角多项式,在这里 α_k, β_k 是某些常数.

用三角多项式逼近有种种特别的方法;这些方法与用代数多项式逼近的对应方法十分简单地关联着.在这些方法中,函数的傅里叶三角级数展开式(§7 专讲傅里叶级数)有特别重要的地位.这级数是以法国数学家傅里叶命名的;上世纪初,傅里叶在研究热传导论时,得到了关于这种级数的一些理论上的结果.然而必须指出,伟大的数学家列昂纳德·欧拉与达尼尔·伯努利在十八世纪中叶已经研究过三角级数.欧拉是在研究天文学方面的问题时、而伯努利是在研究弦振动时用到这种级数.顺便说明,或多或少任意的函数是否可以用三角级数表示这一有原则意义的问题,是由欧拉与伯努利提出的;这个问题直到上世纪中叶才最后得到解决.伯努利已经预料到这问题有肯定的解答,我们以后还要讲到它的解答.

傅里叶级数在物理上有重大的意义,我们将较少注意到这一方面,因为在第六章中已经研究过这个问题.在那一章中,读者可

以认识到一些物理问题，它们自然地导向必须把已给函数展开成为与三角级数相似的非三角级数．我们要考虑所谓正交函数的级数．

傅里叶级数已有两个世纪的历史．因此到现代已经建立了很广泛的、非常精密与深刻的傅里叶级数论，它已成为数学中的一门独立理论，这是不足为奇的．在这一理论中，苏联的数学有主导的成就，而且在一系列主要问题上有空前绝后的成就．莫斯科实变函数论学派（鲁金、柯尔莫果洛夫、门朔夫等人）在这段历史中起着特别重要的作用．

我们还要指出，三角多项式在现代数学中的意义远远不仅限于作为逼近的工具．例如在第十章中，读者可以知道维诺格拉朵夫在数论方面的一些基本结果，这些结果是他用相应作出的三角和（多项式）作为工具而推得的．

§2. 插值多项式

插值多项式作法的特别情形 逼近函数的插值法在实际计算上有广泛的应用．为了把读者引到这方面的问题，我们要考虑下列初等的问题．

设已给在线段 $[x_0, x_2]$ 上的函数 $y=f(x)$，它的图形画在图 1 上．这图形的形状使我们想到一个抛物线的一段．因此如果我们想用简单的函数逼近已给函数；那么自然选取某一、二次多项式

$$P(x)=a_0+a_1x+a_2x^2 \quad (1)$$

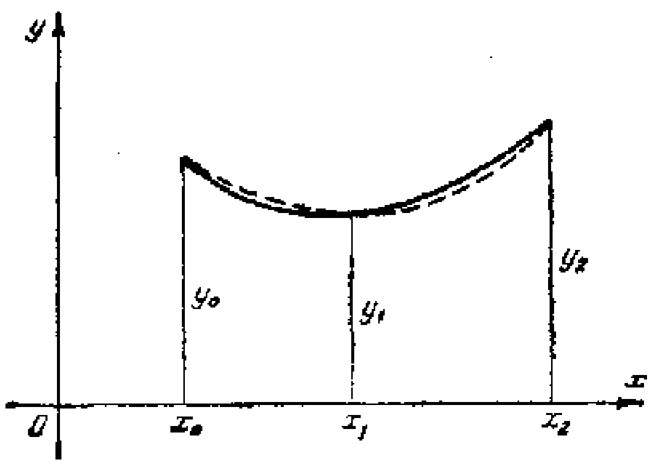


图 1

作为简单的逼近函数，上列多项式的图形是一抛物线．

插值法可叙述如下，在线段$[x_0, x_2]$的内部还给出一内点x_1，与点x_0, x_1, x_2相对应的函数值是$y_0=f(x_0), y_1=f(x_1), y_2=f(x_2)$.

作出形如(1)的多项式，使它在点x_0, x_1, x_2与所考虑的函数(它的图形用虚线画在图1上)有相同的值. 换句话说，需要选取多项式(1)的系数a_0, a_1, a_2使得下列等式成立：

$$P(x_0)=y_0,\ P(x_1)=y_1,\ P(x_2)=y_2. \tag{2}$$

我们指出，函数$f(x)$可能一开始就不能用公式表示出来，例如它可能代表一种经验关系，其图形画在图1上. 解决了它的插值问题，我们就得到表示成解析式、即多项式形状的逼近函数. 如果逼近的准确度可以使人满意，那么求得的多项式比被逼近的函数有这种优点：可以计算中间数值x.

提出的问题可解决如下：作三个方程

$$y_0=a_0+a_1x_0+a_2x_0^2,$$
$$y_1=a_0+a_1x_1+a_2x_1^2,$$
$$y_2=a_0+a_1x_2+a_2x_2^2,$$

从这些方程求出a_0, a_1, a_2, 然后把系数的数值代入等式(1). 但是我们要用有点不同的方法来解决这个问题. 先作出一个二次多项式$Q_0(x)$, 使得它满足三个条件：$Q_0(x_0)=1$, $Q_1(x_1)=0$, $Q_2(x_2)=0$. 由后两个条件可见这多项式必须有$A(x-x_1)(x-x_2)$的形状，而由第一个条件可见$A=\frac{1}{(x_0-x_1)(x_0-x_2)}$因此，所求多项式的形状是

$$Q_0(x)=\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}.$$

同样，多项式

$$Q_1(x)=\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)},\quad Q_2(x)=\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}$$

满足条件

$$Q_1(x_0)=Q_1(x_2)=0,\ Q_1(x_1)=1,$$
$$Q_2(x_0)=Q_2(x_1)=0,\ Q_2(x_2)=1.$$

其次，显然多项式$y_0Q_0(x)$在$x=x_0$时等于y_0，而在$x=x_1$及

$x=x_2$ 时等于零。多项式 $y_1Q_1(x)$ 及 $y_2Q_2(x)$ 也有类似的性质。

由此可见，所求的满足条件(2)的二次插值多项式可以用公式表示如下：

$$P(x)=y_0Q_0(x)+y_1Q_1(x)+y_2Q_2(x)$$
$$=y_0\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}+y_1\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}$$
$$+y_2\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}. \tag{3}$$

我们指出，求得的多项式是解决上述插值问题的唯一的二次多项式。事实上，如果设有另一、二次多项式 $P_1(x)$ 也能解决这个问题，那么差式 $P_1(x)-P(x)$ 也是二次多项式，而且它在 $x=x_0$, x_1, x_2 三点等于零。但由代数我们知道，如果一个二次多项式对于 x 的三个值为零，那么它就恒等于零。因此，多项式 $P(x)$ 与 $P_1(x)$ 恒等。

显然，一般说来，求得的多项式仅仅在点 x_0, x_1, x_2 处与已给的函数有相同的值；而对于 x 的另外的值，它与已给函数有不同的值。

如果取线段 $[x_0, x_2]$ 的中点作为 x_1，令 $x_2-x_1=x_1-x_0=h$，那么公式(3)可以略为简化如下：

$$P(x)=\frac{1}{2h^2}[y_0(x-x_1)(x-x_2)-2y_1(x-x_0)(x-x_2)$$
$$+y_2(x-x_0)(x-x_1)].$$

作为例子，对于正弦 $y=\sin x$（图 2），用与它在点 $x=0, \frac{\pi}{2}$, π 有同值的二次多项式插值。显然，所求的多项式有下列形状：

$$P(x)=\frac{4}{\pi^2}x(\pi-x)\approx\sin x.$$

在两个中间的点比较 $\sin x$ 与 $P(x)$：

$$P\left(\frac{\pi}{4}\right)=0.75,\text{而}\sin\frac{\pi}{4}=\frac{\sqrt{2}}{2}\approx0.71,$$

$$P\left(\frac{\pi}{6}\right)=\frac{10}{18},\text{而}\sin\frac{\pi}{6}=\frac{9}{18}.$$

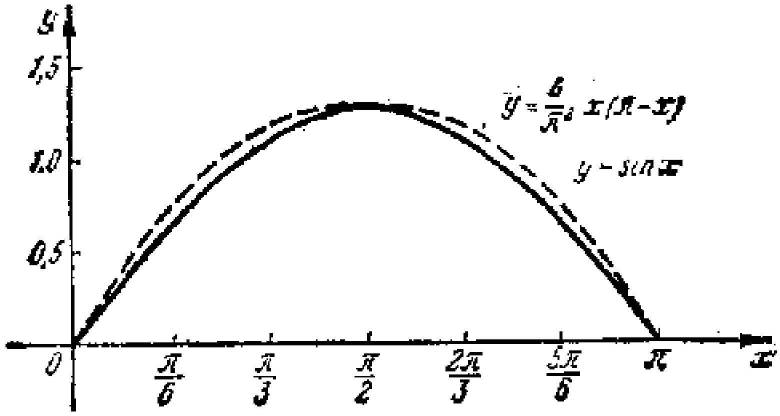


图 2

因此，在线段$[0,\pi]$上逼近$\sin x$，大约准确到0.05[1]．另一方面，在点$\frac{\pi}{2}$的邻域展开$\sin x$为泰勒级数，即得

$$\sin x=\cos\left(\frac{\pi}{2}-x\right)=1-\frac{\left(\frac{\pi}{2}-x\right)^2}{2!}+\frac{\left(\frac{\pi}{2}-x\right)^4}{4!}-\cdots.$$

如果在展开式中取到第二项为止，那么在点$x=0$得到$\sin 0$的逼近：$\sin 0\approx 1-\frac{\pi^2}{8}\approx 0.234$，其误差大于0.2．

我们可以看到，把函数$\sin x$在点$x=\frac{\pi}{2}$的邻域内按泰勒公式展开，取二次多项式逼近$\sin x$，这种方法不如按照插值法、取同次多项式逼近$\sin x$那样令人满意．可是不应当忘记，用泰勒公式可得在$x=\frac{\pi}{2}$的小邻域内很准确的逼近，比用插值法所得在这邻域内的逼近要准确得多．

问题的一般解法　显然，几乎谁也不会决定用二次多项式对图3中所画出的较复杂的函数作插值．不管怎样，谁若是决定这样做，那么显然会得到很坏的逼近，因为任何二次抛物线不能有曲

1）为了完全说明这一断言的理由，应当证明不仅对于$x=\frac{\pi}{4}$及$x=\frac{\pi}{6}$，而且也对于线段$[0,\pi]$上的一切x的值，差$\frac{4x}{\pi^2}(\pi-x)-\sin x$不超过0.05阶的数；但是我们不证明这个结论。

线 $y=f(x)$ 那样多的弯曲. 这时自然要设法用较高次(不低于四次)的多项式对函数作插值.

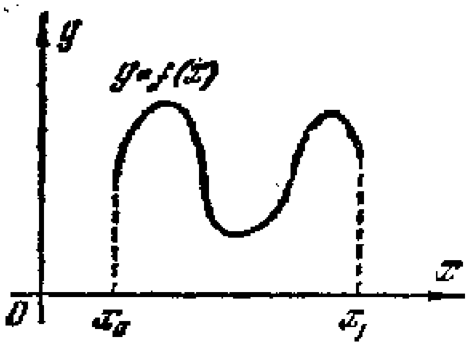


图 3

插值的一般问题是: 作 n 次多项式 $P(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$, 使得它在 $n+1$ 个点 x_0, x_1, x_2, $\cdots$, x_n 与已给函数有相同的值, 也就是说, 使得它满足 $n+1$ 个等式: $P(x_0)=f(x_0)$, $P(x_1)=f(x_1)$, $\cdots$, $P(x_n)=f(x_n)$. 在函数与逼近它的多项式有同值的各点, 称为插值结点. 与对二次多项式同样进行讨论, 不难证明, 所求的多项式可以写成

$$P_n(x)=\sum_{k=0}^{n}\frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}f(x_k),\tag{4}$$

而且这个(n 次)多项式是唯一的. 上面写出的公式称为拉格朗日公式. 这种公式可以写成其它各种形状; 例如牛顿差分公式在实际上有广泛的应用.

插值多项式与生成函数的偏差 插值法是逼近函数的很普遍的方法. 在原则上, 函数不一定要有特殊的性质, 就可以作它的插值; 例如函数不一定要在作逼近的整个区间上有导数. 在这种意义下, 插值法比泰勒公式优越. 有趣的是, 常常有这样的情况: 函数甚至于在一个线段上解析, 而对它不能应用泰勒公式作为逼近的工具. 例如, 我们设想要在线段[−2, 2]上用代数多项式足够好地逼近函数 $\frac{1}{1+x^2}$. 骤然看来, 为此, 自然要想把它在点 $x=0$ 的邻域内展开成泰勒级数

$$\frac{1}{1+x^2}=1-x^2+x^4-x^6+\cdots.$$

可是不难看出, 得到的级数只在区间 $-1<x<1$ 内收敛. 它在线段[−1, 1]以外发散, 因而不可能在整个线段[−2, 2]上逼近函数

$\frac{1}{1+x^2}$. 但是同时在这里完全可以应用插值法.

当然，每次要发生适当选定结点的个数与分布这一问题，使得逼近的误差满足所希望的条件．在函数有足够高阶的导数的情形下，下列古典的结果对逼近可能有的误差这一问题作了答复；我们只叙述这个结果，但不加证明.

如果函数 $f(x)$ 在线段 $[x_0, x_n]$ 上有连续的 $n+1$ 阶导数，那么对于这线段中 x 的任一中间数值，已给函数与以 $x_0<x_1<\cdots<x_n$ 为结点的拉格朗日插值多项式的偏差，可以用公式

$$f(x)-P(x)=\frac{(x-x_0)(x-x_1)\cdots(x-x_n)}{n!}f^{(n+1)}(c)$$

表示出来，在这里 c 是 x_0 与 x_n 之间的一点．这个公式使我们想到泰勒展开式的相应的余项公式；在实质上，前者是后者的推广．这样，如果知道了 $n+1$ 阶导数的绝对值在线段 $[x_0, x_n]$ 上到处不超过数 M，那么对于在这线段上 x 的任一值，逼近误差由下列估计所决定：

$$|f(x)-P_n(x)|\leqslant\frac{|x-x_0|\cdots|x-x_n|}{n!}M.$$

在近代逼近论中还有求插值误差的其它许多方法．这个问题现在已经很好地研究过．同时已经得到了一些完全没有预料到的有趣结果.

例如考虑在线段 $[-1, 1]$ 上所给出的光滑函数 $y=f(x)$，这就是说，它的图形是切线连续变化的连续曲线．取有固定端点 -1 及 1 的线段这一情况，实际不起特殊的作用；以下所讲到的一些事实，在作若干非本质的改变后，对于任意线段 $[a, b]$ 也成立.

现在假定在线段 $[-1, 1]$ 上给出一组 $n+1$ 个点

$$-1\leqslant x_0<x_1<\cdots<x_n\leqslant 1, \tag{5}$$

然后作出一个 n 次多项式 $P(x)=a_0+a_1x+\cdots+a_nx^n$，使得它在上列 $n+1$ 个点与 $f(x)$ 有同值．现在假定在点组(5)中，邻接的点之间有等距离．如果 n 无限增大，那么函数 $f(x)$ 的对应的插值多项式 $P_n(x)$，在愈来愈多的点上与 $f(x)$ 有相同的值；并且可以想到，在

不属于点组(5)的中间点 x，当 $n\to\infty$ 时，差式 $f(x)-P_n(x)$ 趋近于零．在上世纪末已经有过这种想法，然而后来发现，事实远不是这样．原来对于许多光滑(甚至于解析)函数 $f(x)$，在结点 x_k 有等距离的情况下，插值多项式 $P_n(x)$ 当 $n\to\infty$ 时完全不趋近于 $f(x)$．虽然插值多项式的图形在已给结点与 $f(x)$ 相重合，可是付出了这样的代价：当 n 很大时，对于结点之间的中间数值 x，它与 $f(x)$ 的图形相差很远，而且 n 愈大，相差愈远．进一步的研究证明了，如果插值结点在线段的中央分布得较稀，而在端点附近较密，那么至少对于光滑函数，上述现象可以避免．当插值结点 x_k 与车比雪夫多项式 $\cos[(n+1)\arccos x]$ 的零点[1]相重合，亦即当

$$x_k=\cos\frac{2k+1}{2(n+1)}\pi\quad(k=0,\ 1,\ \cdots,\ n)$$

时，在一定的意义下，正是这种分布是插值结点的最好分布．

与这种结点相对应的插值多项式，以车比雪夫命名，并且具备着这样的性质：当 n 无限增大时，它一致收敛于产生它的光滑函数，亦即不仅本身连续、而且有连续一阶导数的函数．这种函数的图形是切线连续变化的连续曲线．在图 4 上，画出了当 $n=5$ 时，车比雪夫多项式的零点的分布．

对于任意不光滑的连续函数，那么情形就要坏些；原来一般说

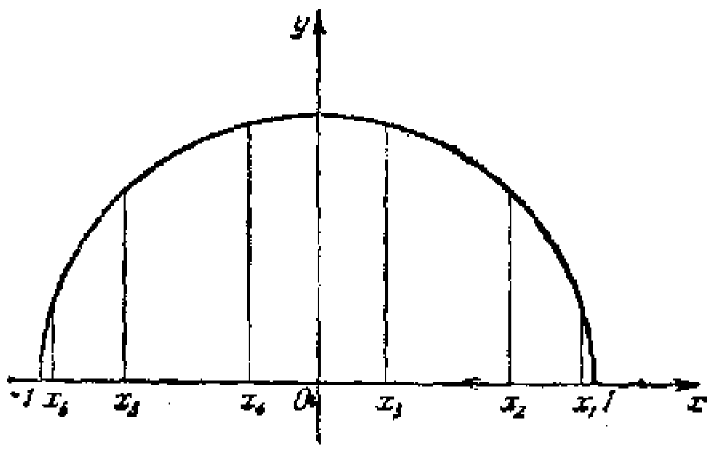


图 4

1) 任何适合 $f(x_k)=0$ 的数值 x_k 称为函数 $f(x)$ 的零点．关于车比雪夫多项式的更详细的讨论，见 §5．

来，不存在着这样的插值结点序列：使得对于任一连续函数，对应的插值过程收敛(法柏尔定理)．换句话说，无论怎样分割线段$[-1, 1]$，得到结点的个数无限增大，总可找到在这线段上连续的函数$f(x)$，使得按这些结点逐步插补它的多项式不收敛于已给函数．如果上世纪中叶的数学家知道了这个事实，一定会认为这是难以置信的．当然，问题在于在连续不光滑的函数中，有些很“坏”的函数，例如有在给出函数的线段上处处没有导数的函数．这就是为什么在这种函数中可以找到一些函数，对于它们给定的插值过程发散．可是对于它们还是可以提出有效的多项式逼近法，这些方法是上述插值法的变形；不过我们不讲这个问题．

最后我们指出，可以不一定用代数多项式对函数作插值．例如在实用方面以及在理论方面已经很好地研究过用三角多项式的插值法。

§3．定积分的逼近

在与积分近似计算有关的问题中，函数插值法有广泛的应用．作为例子，我们举出定积分的一个近似表示公式，即辛卜森公式；这个公式在实用分析中用处特别大．

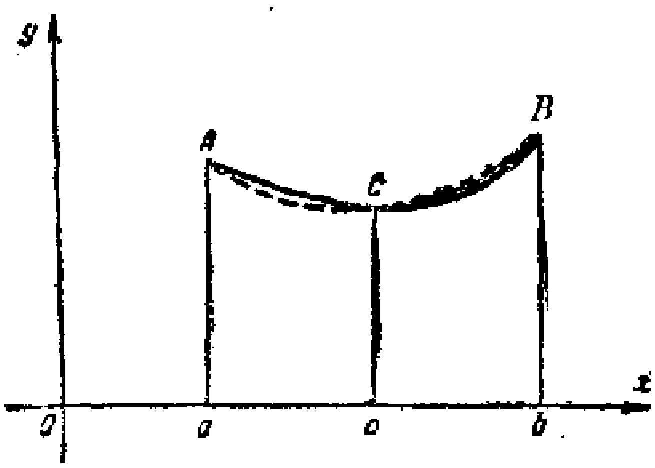


图 5

设要近似地计算函数$f(x)$在线段$[a, b]$上的定积分；这函数的图形见图5．定积分的大小恰好等于曲线梯形$aABb$的面积．设C是图形上以$c=\frac{a+b}{2}$为横坐标的点．过点A、B及C作二次抛物线．由上节可知，这条抛物线是由下列等式所确定的二次多项式的图形：

$$P(x)=\frac{1}{2h^2}[(x-c)(x-b)y_0-2(x-a)(x-b)y_1$$
$$+(x-a)(x-c)y_2],$$

在这里

$$h=\frac{b-a}{2}, y_0=f(a), y_1=f(c), y_2=f(b).$$

应用上节中已用过的术语，可以说二次多项式 $P(x)$ 在以 a、c、b 为横坐标的点插补函数 $f(x)$. 如果函数 $f(x)$ 的图形在线段 $[a, b]$ 上变化得很平滑，而且线段本身不长，那么多项式 $P(x)$ 到处与 $f(x)$ 相差很小；由此可推出这两函数在 $[a, b]$ 上的积分也相差很小. 根据这一点，我们可以把这两积分算作近似相等：

$$\int_a^b f(x)dx\approx\int_a^b P(x)dx,$$

或者也可以把第二个积分算作第一个的近似值. 读者作了简单的计算以后，可以看出

$$\int_a^b (x-c)(x-b)dx=\frac{2}{3}h^3, -\int_a^b (x-a)(x-b)dx=\frac{4}{3}h^3,$$
$$\int_a^b (x-a)(x-c)dx=\frac{2}{3}h^3,$$

从而 $$\int_a^b P(x)dx=\frac{h}{3}[f(a)+4f(c)+f(b)].$$

这样，我们的定积分可用下列近似公式算出：

$$\int_a^b f(x)dx\approx\frac{h}{3}[f(a)+4f(c)+f(b)].$$

这就是辛卜森公式.

作为应用这公式的例子，我们计算 $\sin x$ 在线段 $[0, \pi]$ 上的积分. 这时

$$h=\frac{\pi}{2}, f(a)=\sin 0=0, f(c)=\sin\frac{\pi}{2}=1, f(b)=\sin\pi=0,$$

因而 $\frac{h}{3}[f(a)+4f(c)+f(b)]=\frac{2}{3}\pi=2.09\cdots$. 另一方面，我们可以准确地求出这个积分

$$\int_0^{\pi}\sin x\,dx=-\cos x\Big|_0^{\pi}=2,$$

误差不超过 0.1.

如果把区间 $[0,\pi]$ 分成两个相等的部分，并且对其中每一部分分别应用上述公式，那么就得到

$$\int_0^{\frac{\pi}{2}}\sin x\,dx\approx\frac{\pi}{12}\left[\sin 0+4\sin\frac{\pi}{4}+\sin\frac{\pi}{2}\right]$$
$$=\frac{\pi}{12}\left(4\cdot\frac{\sqrt{2}}{2}+1\right)\approx 1.001,\quad \int_{\frac{\pi}{2}}^{\pi}\sin x\,dx\approx 1.001.$$

于是

$$\int_0^{\pi}\sin x\,dx\approx 2.002;$$

现在误差已更加减小，约为 0.002.

在实用上，为了近似计算函数 $f(x)$ 在线段 $[a,b]$ 上的定积分，把这线段用点 $a=x_0<x_1<\cdots<x_n=b$ 分成偶数 n 个部分，并且先对线段 $[x_0,x_2]$，然后对线段 $[x_2,x_4]$ 等等，逐步应用辛卜森公式. 结果得到下列一般的辛卜森公式：

$$\int_a^b f(x)dx\approx\frac{b-a}{3n}[f(x_0)+4f(x_1)+2f(x_2)+4f(x_3)$$
$$+\cdots+f(x_n)]. \tag{6}$$

我们举出对这公式的古典的估计. 如果函数 $f(x)$ 在线段 $[a,b]$ 上有满足不等式 $|f^{IV}(x)|\leqslant M$ 的四阶导数，那么下列估计式成立：

$$\left|\int_a^b f(x)dx-L(f)\right|\leqslant\frac{M(b-a)^5}{180n^4}. \tag{7}$$

这里用 $L(f)$ 表示公式(6)的右边. 在这种情形下，误差有 n^{-4} 阶[1].

我们也可以把线段 $[a,b]$ 分成 n 个相等的部分，并且把图 6 上画出的那些矩形的面积的和称作积分的近似值. 这时我们就得

1) 如果与 $n=1,2,\cdots$ 有关的一数 α_n 满足等式 $|\alpha_n|<\frac{C}{n^k}$，在这里 C 是与 n 无关的常数，那么我们说它有 n^{-k} 阶.

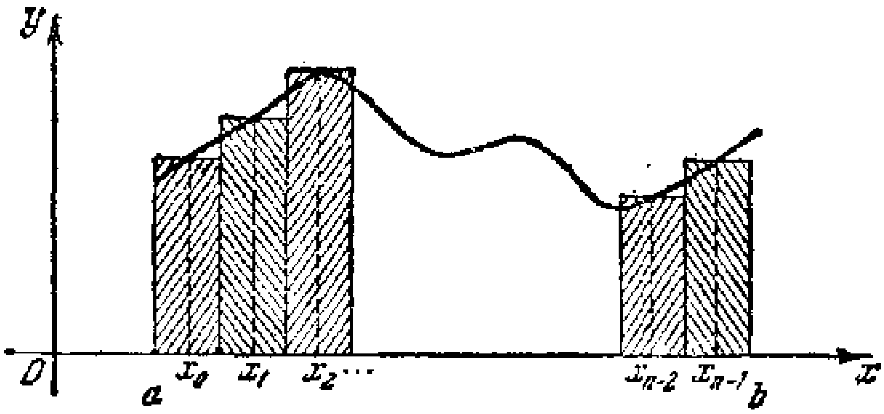


图 6

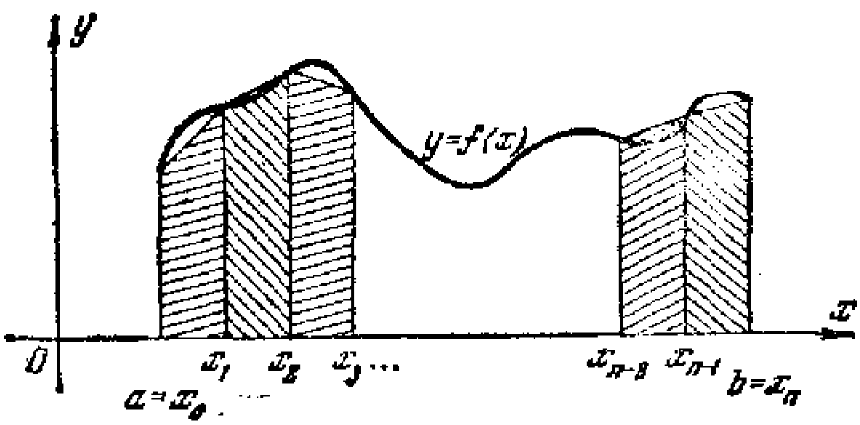


图 7

到矩形近似公式[1]:

$$\int_a^b f(x)dx \approx \frac{b-a}{n}[f(x_0)+f(x_1)+\cdots+f(x_{n-1})]. \qquad (8)$$

可以证明，只要函数在线段$[a, b]$上有二阶导致，这公式的误差的阶是n^{-2}. 我们也可以把图7上画出的那些梯形的面积的和取作积分的近似值，这时得到梯形公式

$$\int_a^b f(x)dx \approx \frac{b-a}{2n}[f(x_0)+2f(x_1)+\cdots+2f(x_{n-1})+f(x_n)]; \qquad (9)$$

如果函数有有界的二阶导数，这公式有误差阶n^{-2}.

通常说来，辛卜森公式比梯形及矩形公式准确．这句话需要解释一下．如果我们只知道函数有一阶导数，那么这三种公式所

1) 在这种情形下，$x_0, x_1, \cdots, x_{n-1}$是区间$[a, b]$中相等的部分的中点，而不象在公式(6)及(9)中那样是分点．

求得的积分近似值都一定有 n^{-1} 阶；在这种情形下，辛卜森公式比矩形及梯形公式并没有实质上的优越性。对于有二阶导数的函数，已可保证用梯形及辛卜森公式所求得的近似值有 n^{-2} 阶。如果函数有三阶及四阶导数，那么对于矩形及梯形公式，误差的阶仍然等于 n^{-2}，而对于辛卜森公式，则分别等于 n^{-3} 及 n^{-4}。但对于辛卜森公式本身，次数 n^{-4} 也是误差的界限，这就是说，对于有高于四阶导数的函数，误差的阶还是等于 n^{-4}。因此如果已经有五阶导数的函数，并且我们希望利用这一情况，求得 n^{-5}阶的近似值，那么就需要作出与辛卜森公式不同的、新的定积分逼近法。为了了解应当怎样作出这种方法，我们说明如下。

容易验证，梯形及矩形公式对于一次多项式是准确的；这就是说，在(9)中代入函数 $A+Bx$，就得到准确的等式，在这里 A 及 B 是常数。在这种意义下，辛卜森公式对于三次多项式 $A+Bx+Cx^2+Dx^3$ 是准确的。问题正是在于这些性质。我们设想把线段 $[a, b]$ 分成 n 等分，对每部分应用对 $m-1$ 次多项式 $A+Bx+\cdots+Fx^{m-1}$ 准确的同一种积分逼近法。于是对于任一有有界 m 阶导数的函数，逼近误差就有 n^{-m} 阶，并且如果这函数不是 $m-1$ 次多项式，那么即令对于有较高阶导数的函数，误差的阶也不能提高。

我们已经讨论到这一问题的重要性：找出尽可能简单的积分的近似方法，使其对已给次数的多项式准确。很久以来，数学家已经注意到这个问题，现在在这方面有许多文献。在这里我们只讲几个古典的结果。

已给函数 $p(x)$。试问应当怎样在线段 $[-1, 1]$ 上安排结点 $x_1, \cdots, x_m$，并且怎样选择数 K，使得无论 $f(x)$ 是任何 m 次多项式，下列等式成立：

$$\int_{-1}^{1} f(x)p(x)dx = K\sum_{1}^{m} f(x_i).$$

原来当 $p(x)=(1-x^2)^{-\frac{1}{2}}$ 时，如果 $K=\frac{\pi}{m}$，而 x_i 是以车比雪夫命名的多项式 $\cos m \arccos x$ 的零点，那么上述问题有肯定的解答。

对于 $p(x)=1$，在 $m=1, 2, \cdots, 7$ 时，车比雪夫给出了这个问题的解答。在 $m=8$ 时，这个问题没有解：可以找到结点，但它们是复数。在 $m=9$ 时，这个问题又有解。然而伯恩斯坦曾经证明，对于任何 $m>9$，问题仍然无解（结点在线段 $[-1, 1]$ 以外）。

应用拉格朗日公式(4)，可以很简单地推出对 n 次多项式准确的求积公式。如果对(4)式的左右两边在线段 $[a, b]$ 上求积分，那么得到

$$\int_a^b P_n(x)dx=\sum_{k=0}^{m} p_k f(x_k), \tag{10}$$

在这里

$$p_k=\int_a^b \frac{(x-x_0)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}dx$$
$$(k=0, 1, \cdots, n).$$

于是等式(10)对于一切 n 次多项式成立，因而求积公式

$$\int_a^b p(x)dx \approx \sum_0^n p_k f(x_k)$$

对于一切 n 次多项式准确。

我们知道，当

$$x_0=a,\ x_1=\frac{a+b}{2},\ x_2=b$$

时，这公式化为辛卜森公式。

我们可以改变结点 $x_k(k=0, 1, \cdots, n)$ 在线段 $[a, b]$ 的范围内的分布。对应于结点的每种分布，有一求积公式。

高斯（上世纪著名的德国数学家）曾经证明，可以安排结点 x_k，使得求积公式不仅对于一切 n 次多项式、而且也对于一切 $2n+1$ 次多项式准确。

利用高斯的结点 x_k 所作出的 $n+1$ 次多项式

$$A_{n+1}(x)=(x-x_0)(x-x_1)\cdots(x-x_n)$$

有巧妙的性质：无论 $P(x)$ 是任何次数小于 $n+1$ 的多项式，下列等式成立：

$$\int_a^b A_{n+1}(x)P(x)dx=0.$$

这样，多项式 $A_{n+1}(x)$ 在线段 $[a, b]$ 上与一切次数不超过 n 的多项式正交。多项式 $A_{n+1}(x)$ 称为(对应于线段 $[a, b]$ 的) **莱让德尔多项式**。

§4. 车比雪夫最好一致逼近的观念

问题的提出 车比雪夫由于希望解决纯粹实用的问题，而得到了最好一致逼近的观念。车比雪夫不仅是上世纪的一位伟大的数学家，建立了一系列现在已经广泛发展了的数学理论的基础，而且他也是当时的一位先进的工程师。特别，车比雪夫对这样的问题很感兴趣：即关于能作出某种已给运动轨线的机械的构造问题。我们现在来说明这些话表示什么。

设在线段 $a \leqslant x \leqslant b$ 上已给曲线 $y=f(x)$。需要造成满足一定技术条件的机械，使得当机械运转时，它上面的一点尽可能准确地画出已给曲线。车比雪夫解决所提出的问题如下。首先他作为工程师来解决这个问题，造成了需要的机械，利用它可以比较约略近似地得到所需的运动轨线。这样，当这暂且还不是最后设计的机械运转时，它上面的一点 A 画出曲线

$$y=\varphi(x), \tag{11}$$

这曲线只与所需的曲线 $y=f(x)$ 大略相似。造出的机械是由一些各别的部分、如小齿轮、各类杠杆等等所构成的。所有这些部分都有一定的大小

$$\alpha_0,\ \alpha_1,\ \alpha_2,\ \cdots,\ \alpha_m, \tag{12}$$

这些数量完全确定了机械，因而也完全确定了曲线(11)。这就是机械以及曲线(11)的参数[1]。因此，曲线(11)不仅与自变数 x 有关，而且也与参数(12)有关。对应于任意一组参数的数值，有一条确定的曲线，其方程可以方便地写成下列形状：

$$y=\varphi(x;\ \alpha_0,\ \alpha_1,\ \cdots,\ \alpha_m). \tag{13}$$

1) 希望知道这类机械设计的详细情形的读者，请参看“车比雪夫的科学遗产”论文集，第二卷，苏联科学院出版社，1945 年版。

在这种情形下，照例说我们已经得到了一族函数(12)，这些函数是在线段 $a\leqslant x\leqslant b$ 上给出、并且与 $m+1$ 个参数(11)有关.

在进一步解决他自己的问题时，车比雪夫已作为纯粹数学家而出现．他十分自然地提出了把数量

$$\|f-\varphi\|=\max_{a\leqslant x\leqslant b}|f(x)-\varphi(x;\alpha_0,\alpha_1,\cdots,\alpha_m)| \qquad (14)$$

算作已给函数 $f(x)$ 与邻近它的函数 $\varphi(x;\alpha_0,\alpha_1,\cdots,\alpha_m)$ 在线段 $a\leqslant x\leqslant b$ 上的偏差的尺度，这个数量等于差式 $f(x)-\varphi(x;\alpha_0,\alpha_1,\cdots,\alpha_m)$ 的绝对值在线段 $a\leqslant x\leqslant b$ 上的极大值(图 8)，显然，这数量是参数 $\alpha_0,\alpha_1,\cdots,\alpha_m$ 的一个函数

$$\|f-\varphi\|=F(\alpha_0,\alpha_1,\cdots,\alpha_m). \qquad (15)$$

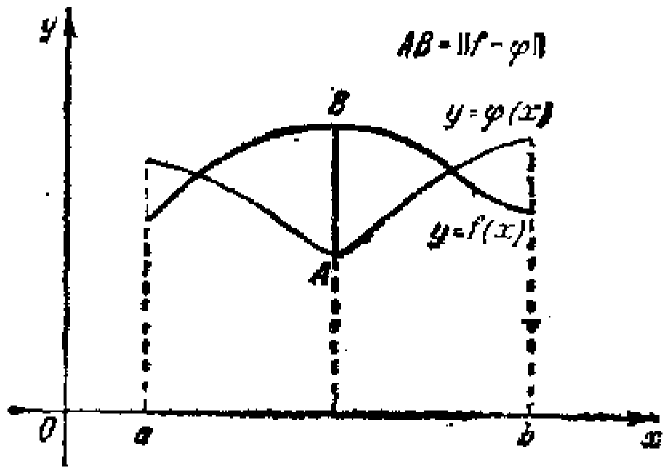


图 8

现在提出寻找使函数(15)取最小值的参数数值这一问题．这些数值确定函数 φ；通常说在所考虑的族(11)内的一切函数中，这样确定的函数 φ 以最好的方式一致逼近已给函数 $y=f(x)$．对于确定 φ 的这些数值，数量 $F(\alpha_0,\alpha_1,\cdots,\alpha_m)$ 称为函数 $f(x)$ 在线段 $[a,b]$ 上用族(13)中的函数 φ 的**最好一致逼近**．通常把它用符号 $E_m(f)$ 来表示．“一致”这个词，特别在外国的文献中，往往被“车比雪夫”这个词所代替．两个词都着重指出了函数邻近的一定的性质．还可以有其它的方法，例如可以说某一族中的一个函数在平均平方的意义下以最好的方式逼近函数 $f(x)$．在 § 8 中要讲到这个问题．

车比雪夫首先揭露了这里考虑的问题所适合的一些规律，并

且在许多情形下，发现了在线段$[a, b]$上以最好的方式逼近函数f的函数φ，具有一种巧妙的性质：对于这个函数φ，差式

$$f(x)-\varphi(x; a_0, a_1, \cdots, a_m)$$

至少在$[a, b]$中的$m+2$个点上，以正负交错的符号达到它的绝对值的极大值(图 9)。

我们在这里不准确叙述使这断语成立的一些条件，有准备的读者可阅读冈察洛夫的一篇论文“函数的最好逼近论”（“车比雪夫的科学遗产”，第一卷）。

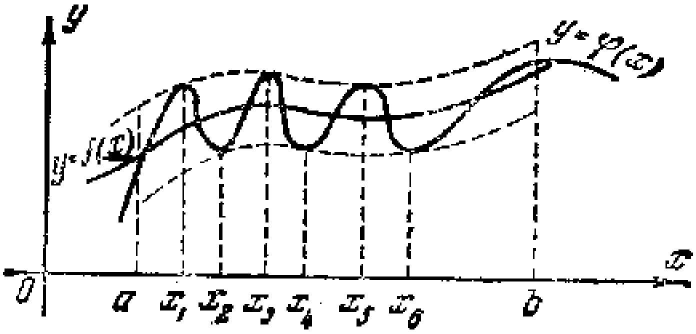


图 9

多项式逼近函数的情形 把上述车比雪夫的研究应用到下列问题上有特别重要的一般理论上的意义：用已给次数n的多项式$P_n(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$逼近在线段$[a, b]$上的任一函数。$n$次多项式$P_n(x)$形成与$n+1$个系数(参数)有关的一个函数族。可以证明，对于多项式完全可以应用车比雪夫的理论；因此，如果要知道在有已给次数n的一切多项式中，给出的多项式$P_n(x)$是否在$[a, b]$上以最好的方式一致逼近于已给函数$f(x)$，那么需要在这线段上找出x的一些数值，使得对于这些数值，函数$|f(x)-P_n(x)|$达到它在$[a, b]$上的极大值L。如果这时从这些数值中可以找出$n+2$个$x_1, x_2, \cdots, x_{n+2}$，使得对于它

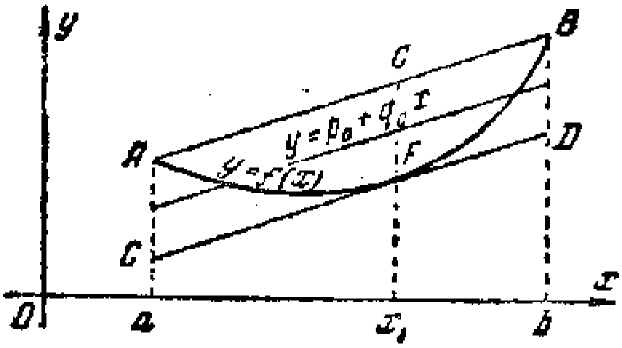


图 10

们差式 $f(x)-P_n(x)$ 依次改变符号：

$$f(x_1)-P_n(x_1)=\pm L,$$
$$f(x_2)-P_n(x_2)=\mp L,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$f(x_{n+2})-Pn(x_{n+2})=\pm(-1)^{n+1}L,$$

那么 $P_n(x)$ 是最好的多项式；否则 $P_n(x)$ 就不是最好的多项式. 例如已给画在图 10 上的函数 $f(x)$. 用一次多项式 $P_1(x)=p+qx$ 最好一致逼近 $f(x)$ 这一问题的解，是一个一次多项式 p_0+q_0x，其图形是一直线，它与弦 AB 平行，并且把这弦以及与它相平行的切线 CD（曲线 $y=f(x)$ 的切线）之间的平行四边形分成相等两部分. 这是因为差式 $f(x)-(p_0+q_0x)$ 的绝对值显然对于数值 $x_0=a$, x_1 以及 $x_2=b$ 达到它的极大值，其中 x_1 是切点 F 的横坐标；而且差式对于上述三数值依次改变符号. 为了避免误会起见，我们指出，这里所讲的是在每点有切线的下凸函数. 在这个例子中，$E_1(f)$ 等于线段 AC 的长的一半，或者等于与 AC 等长的线段 BD 或 GF 的长的一半.

§5. 与零偏差最小的车比雪夫多项式

我们来研究下列问题. 找出在线段 $[-1, 1]$ 上最好一致逼近函数 x^n 的 $n-1$ 次多项式 $P_{n-1}(x)$.

原来所求的多项式满足等式

$$x^n-P_{n-1}(x)=\frac{1}{2^{n-1}}\cos n\arccos x. \tag{16}$$

这个事实可以纯粹形式地从车比雪夫定理推出，只要预先证明了：第一，(16) 的右边是一个 n 次代数多项式，其中 x^n 的系数等于 1；第二，它的绝对值在所考虑的线段 $[-1,1]$ 上 $n+1$ 个点 $x_k=\cos\frac{k\pi}{n}$ $(k=0, 1, \cdots, n)$ 达到极大值 $L=\frac{1}{2^{n-1}}$；第三，它本身在这些点依次改变符号.

用下列推理，可以证明(16)的右边是 n 次多项式，其中 x^n 的系数等于 1.

设对于已给的自然数 n，已经证明了下列等式成立：

$$\cos n \arccos x = 2^{n-1}[x^n - Q_{n-1}(x)],$$

$$-\sqrt{1-x^2}\sin n \arccos x = 2^{n-1}[x^{n+1} - Q_n(x)],$$

在这里 Q_{n-1} 及 Q_n 是次数分别为 $n-1$ 及 n 的代数多项式. 于是对于 $n+1$，类似的等式也成立，因为考虑下列公式就容易证明这一点：

$$\cos(n+1)\arccos x = x\cos n \arccos x - \sqrt{1-x^2}\sin n \arccos x,$$

$$-\sqrt{1-x^2}\sin(n+1)\arccos x = -x\sqrt{1-x^2}\sin n \arccos x + (x^2-1)\cos n \arccos x.$$

而由于

$$\cos \arccos x = x,$$

$$-\sqrt{1-x^2}\sin \arccos x = x^2 - 1,$$

我们的等式对于 $n=1$ 也成立. 在这种情形下，它们对于任何 n 成立.

以首先提出并且解决问题的车比雪夫命名，(16)的右边称为**与零偏差最小的 n 次车比雪夫多项式**. 以下就是这种多项式中的前几个：

$$T_0(x) = 1,$$

$$T_1(x) = x,$$

$$T_2(x) = \frac{1}{2}(2x^2 - 1),$$

$$T_3(x) = \frac{1}{4}(4x^3 - 3x),$$

$$T_4(x) = \frac{1}{8}(8x^4 - 8x^2 + 1),$$

$$T_5(x) = \frac{1}{16}(16x^5 - 20x^3 + 5x).$$

我们已经有机会确认车比雪夫多项式在插值法以及积分逼近法问题中的重要作用. 在这里对插值法作进一步的说明是适宜的.

由于任一连续函数 $f(x)$ 与逼近它的多项式 $P_n(x)$ 的差 $f(x)-P_n(x)$ 在 $n+2$ 点改变符号，于是根据连续函数的性质，$P_n(x)$ 与 $f(x)$ 在线段 $[a, b]$ 上某 $n+1$ 个定点有相同的值，亦即对于选出的一组结点，$P_n(x)$ 是 $f(x)$ 的 n 次插值多项式。

这样，连续函数 $f(x)$ 的最好一致逼近问题，化成会在线段 $[-1, 1]$ 上选择一组插值结点 $x_0, x_1, \cdots, x_n$，使得对应的 n 次插值多项式有可能最小的偏差 $\|f-Q\|=\max\limits_x|f(x)-Q(x)|$ 这一问题．可惜实际找需要的结点常常有很大的困难．通常必须近似地解决这个问题，在这里显出车比雪夫多项式的特殊作用．原来如果恰好取多项式 $\cos(n+1)\arccos x$ 的零点(在这些点，多项式本身等于零)作为插值结点，那么至少对于大的 n，对应的插值多项式与(充分光滑的)函数的一致偏差，以及最好一致逼近函数的多项式与函数的相应偏差，两者相差很小．在一系列重要的特异情形下，“相差很小”这个有点不确定的说法可以用很准确的数量估计来充实，我们在这里不讲这种估计．

现在回到车比雪夫多项式．把它取作下列形状：$T_n(x)=M\cos n\arccos x\ (-1\leqslant x\leqslant 1)$，在这里 M 是一正数．显然，这多项式的绝对值在线段 $[-1, 1]$ 上到处不超过数 M．它的导数等于 $T'_n(x)=-\dfrac{nM\sin n\arccos x}{\sqrt{1-x^2}}$，因而在线段 $[-1, 1]$ 上满足不等式 $|T'_n(x)|\leqslant-\dfrac{nM}{\sqrt{1-x^2}}$．原来对于在线段 $[-1, 1]$ 上绝对值不超过 M 的一切 n 次多项式，上列不等式正确；这就是说，对于在线段 $[-1, 1]$ 上的任何这种多项式的导数下列不等式成立：

$$|P'_n(x)|\leqslant\frac{nM}{\sqrt{1-x^2}}.$$

这个不等式应当属于马尔科夫，因为由他的结果可以直接推出这个于等式，而且甚至他的结果还更深远一点．马尔科夫本人是在研究门捷列夫对他所提出的问题时得到他自己的结果的．

在 1912 年，对于三角多项式，伯恩斯坦进一步得到了以他命名

的、与这不等式相类似的结果，并且利用这些不等式，他首先发现了如果知道函数的最好逼近递减的阶，那么可以推出函数有可微分性．关于可微函数的这种类型的结果，我们要在§6及§7中讲到．

§6．魏尔斯特拉斯定理．函数的最好逼近与它的微分性质

魏尔斯特拉斯定理 如果把§4中所给出的函数的最好逼近一般定义应用到用 n 次代数多项式逼近函数的情形，那么我们就得到下列定义．已给在线段 $[a, b]$ 上的函数 $f(x)$．表达式

$$\max_{a\leqslant x\leqslant b}|f(x)-P_n(x)|=\|f-P_n\|$$

关于一切 n 次多项式 $P_n(x)$ 的极小值等于(非负的)数 $E_n(f)$．那么 $E_n(f)$ 称为函数 $f(x)$ 在线段 $[a, b]$ 用 n 次多项式的最好一致逼近．不论我们是否会找出一个多项式的准确表达式最好一致逼近已给函数，寻求数量 $E_n(f)$ 的更加准确的数值具有重大的实际及理论意义．事实上，如果我们要以达到已给数量 δ 的准确度用多项式逼近函数 f，换句话说，要使得对于已给线段上的一切 x，

$$|f(x)-P_n(x)|\leqslant\delta, \tag{17}$$

那么对于满足 $E_n(f)>\delta$ 的 n，从 n 次多项式中选取逼近多项式就没有什么意义，因为对于这个 n，不存在满足(17)的多项式 P_n，不论 P_n 如何取．另一方面，如果我们知道了 $E_n(f)<\delta$，那么对于这种 n，设法找出以准确度 δ 逼近 $f(x)$ 的多项式 $P_n(x)$ 就有意义，因为这种多项式显然存在．对于不同各类函数的最好逼近的性质，已经有过精细而深刻的研究．首先我们指出下列重要事实．

如果函数 $f(x)$ 在线段 $[a, b]$ 上连续，那么它的最好逼近 $E_n(f)$ 当 n 无限增大时趋近于零．

这是魏尔斯特拉斯在上世纪末证明的一个定理．它有很大的意义，因为它肯定了在原则上：可以用某一多项式以任何预先指

定的准确度逼近任一连续函数．由于这种情况，一切任意次多项式的集合在一定程度上对在线段上所给出的一切连续函数的集合的关系，也象一切有理数的集合 R 对一切实(有理及无理)数的集合 H 的关系．事实上，无论 α 是什么无理数，并且无论 ε 是怎样小的正数，总可找出有理数 r 满足不等式 $|\alpha-r|<\varepsilon$．另一方面，如果 $f(x)$ 是在 $[a, b]$ 上的连续函数，ε 是任一小正数，那么由魏尔斯特拉斯定理，有一代数多项式 $P_n(x)$ 存在，使得对于在 $[a, b]$ 上的一切 x，$|f(x)-P_n(x)|<\varepsilon$ 成立．原来由上述定理，连续函数的最好逼近 $E_n(f)$ 当 $n\to\infty$ 时趋近于零．

可以把魏尔斯特拉斯定理用图形解说如下．设已给确定在线段 $[a, b]$ 上的任一连续函数 $f(x)$ 的图形(参看图 9)．给出任一小正数 ε，围绕图形作出宽度为 2ε 的带形，使得图形通过带形的中央．于是总可选出次数足够高的代数多项式

$$P_n(x)=a_0+a_1x+\cdots+a_nx^n,$$

使得它的图形完全落在所取的带形以内．

我们还作下列说明．设与上面一样，$f(x)$ 是在 $[a, b]$ 上的任一连续函数，$P_n(x)$ $(n=1, 2, \cdots)$ 是它的最好一致逼近的多项式．不难看出，函数 $f(x)$ 可以表示成为在 $[a, b]$ 上一致收敛于它的级数 $f(x)=P_1(x)+[P_2(x)-P_1(x)]+[P_3(x)-P_2(x)]+\cdots$ 这是由于级数的前 n 项的和等于 $P_n(x)$，并且 $\max\limits_{a\leqslant x\leqslant b}|f(x)-P_n(x)|=E_n(f)$，而 $E_n(f)$ 当 $n\to\infty$ 时趋近于零．

结果我们得到魏尔斯特拉斯定理的一种新表述法：

任何在线段 $[a, b]$ 上连续的函数可以表示成为一致收敛于它的代数多项式级数．

这个结果有重大的原则上的意义．它肯定了可以把随意给出的(例如用图形)任一连续函数表示成为解析式的形状．(所谓解析表示就是指或者初等函数，或者由初等函数序列取极限所得的函数．)在历史上，这个结果最后破坏了在数学上几乎到上世纪中叶还一直存在着的关于解析式的种种概念．我们说"最后"，因为

这种类型的一系列特别是关于傅里叶级数的结果，出现在魏尔斯特拉斯定理以前. 在得到这结果以前，人们以为解析函数确定一种特别好的规律性，而这种规律性就是解析函数的特征. 当时通常以为不言而喻，解析式是无穷可微分的，并且甚至可以展开成为幂级数，这些概念是不正确的. 函数可能在整个区间上完全没有导数，而它还是可以用解析式表示出来.

就方法论的观点说来，这些发现的意义在于，从它们可以清楚地觉察到，在原则上数学有可能用它本身的方法来研究比原先想到的更广泛得不可限量的一类规律.

现在已经知道魏尔斯特拉斯定理有许多不同的证法. 其中大部分可以归结为：对于已给的连续函数 f，提出某种步骤来求得一个多项式序列，使得这些多项式当其次数无限增大时对 f 一致逼近，下面就是很简单地作出的一个多项式：

$$B_n(x)=\sum_{k=0}^{n}C_n^k x^k(1-x)^{n-k}f\left(\frac{k}{n}\right),$$

它可以用来逼近在线段 $[0, 1]$ 上连续的函数 $f(x)$. 这个多项式称为伯恩斯坦多项式. 当 n 无限增大时，它在线段 $[0, 1]$ 上一致收敛于产生它的连续函数[1]. 在这里 C_n^k 是从 n 个元素中取 k 个的组合的个数.

我们指出，对复数域也有与魏尔斯特拉斯定理相类似的定理. 在这方面，全面的结果是由拉夫伦捷夫、凯尔迪什以及梅尔格莱得到的.

函数的最好逼近的阶与它的微分性质的关系 我们还指出下列结果. 如果函数 $f(x)$ 在线段 $[a, b]$ 上有 r 阶导数 $f^{(r)}(x)$，其绝对值不超过数 K，那么这函数的最好逼近 $E_n(f)$ 满足不等式

$$E_n(f)\leqslant\frac{c_r K}{n^r}, \tag{18}$$

在这里 c_r 是只与 r 有关的常数（杰克逊定理）. 由不等式(18)可见，

1) 必须说明，虽然伯恩斯坦多项式很简单，可是在实用上很少用到它. 这是因为它收敛得很慢，即令函数有很好的微分性质时还是这样.

当 n 无限增大时，函数有愈高阶数的导数，$E_n(f)$ 趋近于零愈快．这样，函数愈好（愈光滑），它的最好逼近趋近于零愈快．如伯恩斯坦所曾经证明的，在一定的意义下，逆定理也成立．

解析函数比可微分函数还要更好．伯恩斯坦曾经证明，这种函数的最好逼近满足不等式

$$E_n(f) \leqslant cq^n, \tag{19}$$

在这里 c 及 q 是与 f 有关的常数，而且 $0<q<1$，这就是说，$E_n(f)$ 比某一几何级数趋近于零更快．他也证明了：相反，由不等式(19)可推出函数 f 在 $[a, b]$ 上解析．

我们已经引进了在本世纪初所得到的几个最重要的结果，其特征是，它们在很大的程度上，确定了现代函数逼近论的方向．这些结果对于实际估计逼近的意义可以用下列例子加以解释．

如 $Q_n(x)$ 是在线段 $[-1, 1]$ 上按 $(n+1)$ 个结点插补函数 $f(x)$ 的一个 n 次多项式，而且取车比雪夫多项式 $\cos(n+1)\arccos x$ 的零点作为结点，那么在线段 $[-1, 1]$ 上，下列不等式成立：$|f(x)-Q_n(x)| \leqslant c\ln n\, E_n(f)$，其中 c 是与 n 无关的常数，而 $E_n(f)$ 是函数 f 在 $[-1, 1]$ 上的最好逼近．在上一不等式中，按照函数的光滑性，$E_n(f)$ 可以根据(18)或(19)用比它大的数来代替，于是得到用上述插值多项式的好的逼近估计．因为当 n 增大时，$\ln n$ 趋近于无穷大很慢，所以在这种情形下，估计的阶与 $E_n(f)$ 趋近于零的阶相差很小．

用车比雪夫结点的插值的优点在于，对于其它结点，在对应的不等式中，因子 $c\ln n$ 要用增长得更快的因子来代替；在等距离基点的情形，因子特别大．

§7. 傅里叶级数

傅里叶级数的产生 傅里叶级数是在研究某些物理现象、首先是在研究弹性介质的小振动时产生的．普通音乐弦的振动可以作为这种振动的有特征的例子．在历史上，正是振动弦的研究引

导出傅里叶级数，并且确定了其理论的方向．

考虑(图 11) 一根拉紧的弦，其两端固定在轴 Ox 上 $x=0$ 及 $x=l$ 两点．如果使弦离开平衡位置，那么它就要振动。



图 11

我们要研究弦上有横坐标 x_0 的一个定点．它在垂直方向与平衡位置的偏差是时间的一个函数 $\varphi(t)$．原来总可使弦离开平衡，并且给出它在开始时 $t=0$ 的速度，使得结果所要研究的点在垂直方向作由函数

$$\varphi=\varphi(t)=A\cos\alpha kt+B\sin\alpha kt \tag{20}$$

所确定的调和振动，在这里 α 是完全确定的常数，它只与所取的弦的物理性质(密度、张力、长度)有关，k 是已给的任一自然数，A 与 B 是常数．

我们指出，这种讨论是对弦作小振动的情形进行的．因此我们可以近似地把每点 x_0 算作只在垂直方向振动；而略去在水平方向的移动[1)]．我们也把该振动时所产生的摩擦也算作小到可以略去不计．这样，我们近似地把振动算作是不减弱的．

对应于某种调和振动的函数(20) 是有周期 $\frac{2\pi}{\alpha k}$ 的周期函数．由这种函数所确定的振动规律，并不是弦上定点 x_0 可能振动的一切规律．试验以及伴随着它的理论证明，点 x_0 可作某种完全任意

1) 关于这个问题的研究与弦振动微分方程

$$\frac{\partial^2u}{\partial t^2}=a^2\frac{\partial^2u}{\partial x^2}\left(a=\frac{1}{\pi}\alpha\right)$$

有直接的关系；在第六章中，第 57 页上，已经讲过这个问题。

的振动，应当把它看作是由形如(20)的一些完全确定的调和振动相叠加而得．比较简单的规律可由几个这样的振动相加而得，这就是说，它们可以写成形如

$$\varphi(t)=A_0+\sum_{k=1}^{n}(A_k\cos\alpha kt+B_k\sin\alpha kt)$$

的函数，在这里 A_k 及 B_k 是对应的常数．这些函数称为三角多项式．在更复杂的情况下，已给的振动规律必须看作是由无穷个形如(20)的振动相叠加而得，这无穷个振动对应于数 $k=1, 2, 3, \cdots$，并且对应于用应有的方法选出的、与指标 k 有关的某些常数 A_k 及 B_k．这样，我们知道必须把表示点 x_0 的任一振动规律的函数 $\varphi(t)$ $\left(有周期\ \dfrac{2\pi}{\alpha}\right)$，表示成为级数

$$\varphi(t)=A_0+\sum_{k=1}^{\infty}(A_k\cos\alpha kt+B_k\sin\alpha kt) \tag{21}$$

的形状．

在物理中其它一些情况下，不一定表示振动规律的已给函数关系，根据物理上的理由，可自然地看作形如(21)的无穷三角级数的和．例如在振动弦的问题中就有这种情况出现．我们使在试验开始时，弦有例如画在图12上的一定的开始形状．只要知道了确定这形状的函数 $f(x)$ 的三角级数展开式 $f(x)=\sum\limits_{1}^{\infty}a_k\sin\dfrac{k\pi x}{l}$ [它是级数(21)的特殊情形]，我们就容易得到弦振动的准确规律．

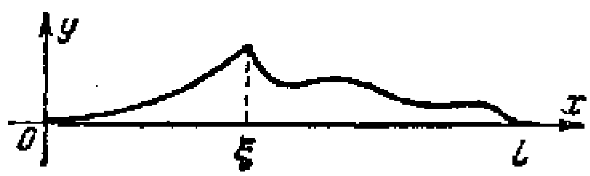


图 12

函数的三角级数展开式 与以上所讲的相关，有一个重要的原则性的问题：怎样的有周期 $\dfrac{2\pi}{\alpha}$ 的函数可以表示成为形如(21)的三角级数的和？在十八世纪，欧拉和伯努利在伯努利关于振动弦的研究中，已经提出这个问题，而且伯努利是站在物理对他所提示的观点上；根据他的观点，应当认为很广泛的一类连续函数，其

中特别包含用手所画出的图形，都可以展开成为三角级数的形状．这种观点受到与伯努利同时的许多数学家的剧烈反对．他们坚持当时存在着的函数概念，根据这些概念，只要函数可用解析式(三角级数就是解析式)表示出来，那么它一定有好的微分性质．但在图 12 上所画出的函数在点 ξ 甚至没有导数；在这种情形下，是否可以对应于已给函数，作出在整个线段 $[0, l]$ 上确定它的解析式？

现在我们知道，伯努利的物理观点恰好是正确的．可是为了使它胜利，还需要整整一个世纪，因为要完全解决这些问题，首先必须明确数学分析中的一些基本概念，象极限以及相关的级数和的概念．

在 1807—1822 年，法国数学家傅里叶进行了证实物理观点的基本数学研究，不过他的研究还是以对于分析的基本概念的旧解释作为基础的．

最后，在 1829 年，德国数学家狄里赫雷以近代数学中所要求的严格性，证明了以 $\frac{2\pi}{\alpha}$ 为周期[1]、并且在一个周期上有有限个极大值及极小值的任何连续函数，可以展开成为一致收敛于它的傅里叶三角级数[2]．

在图 13 上画出了满足狄里赫雷条件的函数．这个连续的图形以 2π 为周期而周期地重复出现，它在周期 $0 \leqslant x \leqslant 2\pi$ 上有一极大点与一极小点

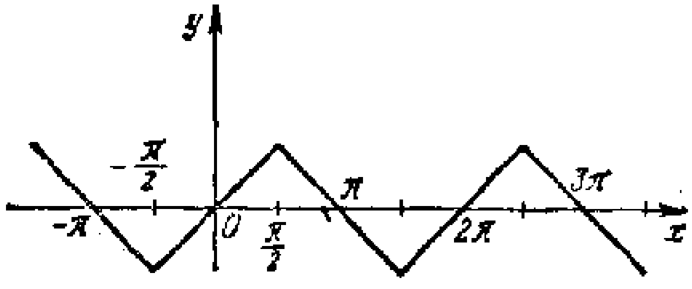


图 13

1) 如果函数 $f(x)$ 满足等式 $f(x+w)=f(x)$，那么我们说它有周期 w．

2) 事实上，狄里赫雷定理也包括了某一类不连续函数，亦即所谓有界变差函数．当然，对于不连续函数，对应的级数不一致收敛．

傅里叶系数 以后我们考虑以 2π 为周期的函数，取 2π 为周期在计算上可以略为简化. 考虑以 2π 为周期、并且满足狄里赫雷条件的任一连续函数. 根据狄里赫雷条件，可以把这函数展开成一致收敛于它的三角级数，

$$f(x)=\frac{a_0}{2}+\sum_{k=1}^{\infty}(a_k\cos kx+b_k\sin kx). \tag{22}$$

我们用 $\frac{a_0}{2}$、而不用 a_0 表示级数的第一项，这并没有原则上的意义；可是如我们将要看到的，这样可得到纯粹在计算上的方便.

我们提出这个问题：计算已给函数 $f(x)$ 的级数的系数 a_k 及 b_k. 为此，我们注意到下列等式成立：

$$\begin{aligned}
&\int_{-\pi}^{\pi}\cos kx\cos lx\,dx=0\quad(k\neq l;\ k,\ l=0,\ 1,\ \cdots),\\
&\int_{-\pi}^{\pi}\sin kx\sin lx\,dx=0\quad(k\neq l;\ k,\ l=0,\ 1,\ \cdots),\\
&\int_{-\pi}^{\pi}\sin kx\cos lx\,dx=0\quad(k,\ l=0,\ 1,\ 2,\ \cdots),\\
&\int_{-\pi}^{\pi}\cos^2 kx\,dx=\pi\quad(k=1,\ 2,\ \cdots),\\
&\int_{-\pi}^{\pi}\sin^2 kx\,dx=\pi\quad(k=1,\ 2,\ \cdots);
\end{aligned}\tag{23}$$

我们建议读者自己证实这些等式.（把不同的三角函数的乘积化成和及差，把三角函数的平方用对应的倍角的三角函数表示出来，就容易算出上列各积分）. 上列等式表明，序列 $1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \cdots$ 中任意两个不同函数的乘积在一个周期上所取的积分等于零(这就是三角函数的所谓正交性). 另一方面，这序列中每个函数的平方的积分等于 π. 但第一个恒等于 1 的函数为例外. 它在一个周期上的积分等于 2π. 由此可见，把级数(22)中的第一个数表示成为 $\frac{a_0}{2}$ 的形状是适当的.

现在可以容易解决我们的问题. 为了计算系数 a_m，先用 $\cos mx$ 乘(22)的左边以及右边的级数的每一项，然后对左边及右边在周

期 2π 上逐项积分；因为乘了 $\cos mx$ 后所得到的级数一致收敛，所以上述计算是合理的．根据等式(23)，右边除了对应于 $\cos mx$ 的积分以外，一切积分都等于零，因此显然

$$\int_{-\pi}^{\pi} f(x)\cos mx\,dx = a_m\pi,$$

从而

$$a_m = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos mx\,dx \quad (m=0,\ 1,\ 2,\ \cdots). \tag{24}$$

同样，用 $\sin mx$ 乘(22)的左边及右边，然后在周期上积分，我们就得到系数的表达式

$$b_m = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin mx\,dx \quad (m=1,\ 2,\ \cdots), \tag{25}$$

于是解决了我们的问题．用公式(24)及(25)算出的数 a_m 及 b_m 称为函数 $f(x)$ 的**傅里叶系数**．

取画在图 13 上的、以 2π 为周期的函数 $f(x)$ 为例．显然，这函数连续，并且满足狄里赫雷条件，因此它的傅里叶级数一致收敛于它．

也容易看出，这函数满足条件 $f(-x)=-f(x)$．显然，函数 $F_1(x)=f(x)\cos mx$ 也满足这个条件．这表示 $F_1(x)$ 的图形关于坐标的原点为对称．根据几何上的理由，显然 $\int_{-\pi}^{\pi} f(x)\,dx=0$，从而 $a_m=0\,(m=0,1,2,\cdots)$．其次，不难明了函数 $F_2(x)=f(x)\sin mx$ 有关于 Oy 轴为对称的图形，从而

$$b_m = \frac{1}{\pi}\int_{-\pi}^{\pi} F_2(x)\,dx = \frac{2}{\pi}\int_0^{\pi} F_2(x)\,dx.$$

但此外，对于偶数 m，$F_2(x)$ 关于轴 Ox 上线段 $[0,\ \pi]$ 的中点 $\frac{\pi}{2}$ 为对称，因此对于偶数 m，$b_m=0$．对于奇数 $m=2l+1\,(l=0,\ 1,\ 2,\ \cdots)$，$F_2(x)$ 的图形关于直线 $x=\frac{\pi}{2}$ 为对称．从而

$$b_{2l+1} = \frac{4}{\pi}\int_0^{\frac{\pi}{2}} F_2(x)\,dx.$$

但由图形可见，在$\left[0,\frac{\pi}{2}\right]$这一部分，简单地有 $f(x)=x$，因而应用分部积分法就得到

$$b_{2l+1}=\frac{4}{\pi}\int_0^{\frac{\pi}{2}}x\sin(2l+1)x\,dx=\frac{4(-1)^l}{\pi(2l+1)^2},$$

从而
$$f(x)=\frac{4}{\pi}\sum_{l=1}^{\infty}\frac{(-1)^l\sin(2l+1)x}{(2l+1)^2}.$$

我们得到了函数的傅里叶级数展开式.

傅里叶部分和对生成函数的收敛性 在应用中，对有周期 2π 的函数 $f(x)$，通常取它的傅里叶级数的前 n 项的和

$$S_n=\frac{a_0}{2}+\sum_{k=1}^{n}(a_k\cos kx+b_k\sin kx)$$

作为函数的逼近，于是就产生了逼近误差的问题. 如果对于一切 x，以 2π 为周期的函数 $f(x)$有满足不等式 $|f^{(r)}(x)|\leqslant K$ 的 r 阶导数 $f^{(r)}(x)$，那么逼近误差可以用下列估计表示出来：

$$|f(x)-S_n(x)|\leqslant\frac{c_r K\ln n}{n^r},$$

在这里 c_r 只是与 r 有关的某一常数. 我们看到，当 n 无限增大时，误差趋近于零；而且函数有愈高阶的导数，误差趋近于零就愈快.

对于在整个实轴上解析的函数，估计还要更好，它可用下列不等式表示出来：

$$|f(x)-S_n(x)|<cq^n, \tag{26}$$

在这里 c 及 q 是与 f 有关的正的常数，而且 $q<1$. 巧妙的是：反过来说，如果对于某一函数，不等式(26)成立，那么这函数是解析的. 发现这个事实是本世纪初的成就；在一定的意义下，它调和了伯努利与他同时的数学家的分歧意见. 现在我们可以说：如果函数可以展开成收敛于它的傅里叶级数，那么由此远不能推出这函数是解析的；如果它与它的傅里叶级数的前 n 项和的偏差递减得比某一递减级数的项更快，那么这函数显然是解析的.

将函数用傅里叶和的逼近的估计，以及用三角多项式的最好逼近的对应估计加以比较，可以看出在光滑函数的情形，用傅里叶

和可得很好的逼近；它接近于最好逼近．然而在不光滑连续函数的情形，结果就要坏些：例如在这些函数中有一些函数，其傅里叶级数在一切有理点所构成的集上发散．

顺便指出，在傅里叶级数论中，还有久已提出，但是至今没有解决的问题；现叙述如下．需要对下列问题作出确定的答复：是否有连续周期函数 $f(x)$，其傅里叶和当 $n\to\infty$ 时对于一切 x 都不收敛于这函数？在这方面，最强最好的结果是柯尔莫果洛夫得到的；他在 1926 年证明了有勒贝格可积的周期函数，其傅里叶和在任一点都不收敛．但可积函数可能不连续；特别，柯尔莫果洛夫所作的函数就是这样．因此这个问题还有待于最后解决．

为了保证任一连续周期函数用三角多项式的逼近，现在应用了傅里叶级数的所谓求和法．代替傅里叶和，我们把它们的某些变形取作用三角多项式的逼近函数．匈牙利数学家费耶尔提出了一种很简单的方法．按照这种方法，对于连续周期函数 $f(x)$，完全形式地作出它的傅里叶级数(可能不收敛)，然后作出这级数的前 n 个部分和的算术平均

$$\sigma_n(x)=\frac{S_0(x)+S_1(x)+\cdots+S_n(x)}{n}. \tag{27}$$

这就是已给函数 $f(x)$ 的所谓 n 阶费耶尔和．费耶尔证明了这种和当 $n\to\infty$ 时一致收敛于 $f(x)$．

§8．在平均平方意义下的逼近

我们回到弦振动的问题．设在某一瞬间 t_0，弦有形状等于 $f(x)$．可以证明，弦的势能、亦即弦从提起的位置回到平衡位置所作的功，等于(当弦的偏差很小时)积分 $W=\int_0^l f'^2(x)dx$，这个结果至少准确到只差一个常数因子．现在设想我们需要用另一函数 $\varphi(x)$ 逼近函数 $f(x)$．与上述弦同时，我们还考虑形状由函数 $\varphi(x)$ 所确定的弦，以及由函数 $f(x)-\varphi(x)$ 所确定的第三条弦．可以证

明，如果第三条弦的能

$$\int_0^l [f'(x)-\varphi'(x)]^2 dx \tag{28}$$

很小，那么前两条弦的能的差更小[1]. 因此，如果第二条弦的能与第一条相差很小对于我们重要，那么必须设法找出函数 $\varphi'(x)$，使得对于它积分(28)尽可能地小. 我们自然地得到了在平均平方的意义下逼近函数[在已给情形下是 $f'(x)$]的问题.

现在讲这问题在一般的形状下是怎样提出的. 在线段 $[a, b]$ 上给出函数 $F(x)$，此外，还给出不仅与 x 有关、而且也与参数 $\alpha_0, \alpha_1, \cdots, \alpha_n$ 有关的函数

$$\Phi(x; \alpha_0, \alpha_1, \cdots, \alpha_n), \tag{29}$$

让这些参数尽可能变化，需要从其中选出参数，使得积分

$$\int_a^b [F(x)-\Phi(x; \alpha_0, \alpha_1, \cdots, \alpha_n)]^2 dx \tag{30}$$

取可能取的最小值. 这个问题在思想上与车比雪夫问题很相似. 在这里讲到的也是函数 $F(x)$ 用函数族(29)的最好逼近，不过是在平均平方意义下的最好逼近. 现在对线段 $[a, b]$ 上的 x 的一切数值，差式 $F-\Phi$ 都很小，对于我们并不重要；在整个线段上的一小部分，差式 $F-\Phi$ 甚至可能很大，只要积分(20)很小就可以了；例如画在图14上的两个图形就是这样. 数量(30)很小就是说平均在线段 $[a, b]$ 的大部分上，函数 F 与 Φ 接近[2]. 关于在实用上选取

---

1) 事实上，如果

$$\int_0^l f'^2 dx \leqslant M^2, \text{并且} \int_0^l \varphi'^2 dx \leqslant M^2,$$

那么

$$\left|\int_0^l f'^2 dx-\int_0^l \varphi'^2 dx\right| \leqslant \left(\sqrt{\int_0^l f'^2 dx}+\sqrt{\int_0^l \varphi'^2 dx}\right)\left|\sqrt{\int_0^l f'^2 dx}-\sqrt{\int_0^l \varphi'^2 dx}\right|$$

$$\leqslant 2M\sqrt{\int_0^l (f'-\varphi')^2 dx}$$

[例如可参看阿赫叶惹尔:"逼近论讲义"第一章，第6页上的闵科夫斯基不等式]，科学出版社1957年版.

2) 在第十九章中(第三卷)，我们将要看到函数在平均平方意义下的近旁，与在通常空间中点的距离之间，有深刻的相似的地方.

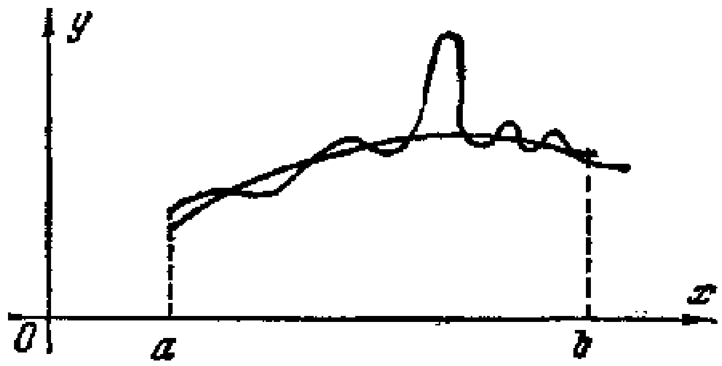


图 14

一种或另一种方法，一切依照我们所提出的目的而定．在刚才所考虑过的弦的例子中，自然要在平均平方的意义下逼近函数 $f'(x)$．另一方面，车比雪夫在解决他的有关机械构造的问题时，平均平方的方法不可能使他满意；因为如果设计的机械零件的表面即使在最小一部分上超出确定的差异范围以外，那么这种零件显然不适用，这样一个凸出的部分损害了整个机械．因此相应于他所提出的实际问题，车比雪夫必须研究一种新的数学方法．

必须说明，就计算观点说来，平均平方法更容易进行些，因为它归结为应用一般分析中很好地研究过的方法。

我们考虑下列有特征的问题作为例子。

需要用形如

$$\sum_{1}^{n} \alpha_k \varphi_k(x)$$

的和式在线段 $[a, b]$ 上按照平均平方的意义最好地逼近已给连续函数 $f(x)$，在这里 α_k 是常数，函数 $\varphi_k(x)$ 是连续的，并且成一规格化正交系．

以上最后一句话就是说，下列等式成立：

$$\int_a^b \varphi_k \varphi_l \, dx = 0, \ k \neq l \ \ (k, l = 1, 2, \cdots, n).$$

$$\int_a^b \varphi_k^2 \, dx = 1 \ \ (k = 1, 2, \cdots, n).$$

考虑数 $a_k = \int_a^b f(x)\varphi_k(x)\,dx \ \ (k = 1, 2, \cdots, n)$，数 a_k 称为函数 f 关于 φ_k 的傅里叶系数。

对于任意的系数 α_k，根据正交及规格化的性质，下列等式成立：

$$\int_a^b\left(f-\sum_1^n \alpha_k\varphi_k\right)^2 dx$$

$$=\int_a^b f^2dx+\sum_1^n \alpha_k^2-2\sum_1^n \alpha_k\, a_k$$

$$=\left[\int_a^b f^2dx-\sum_1^n \alpha_k^2\right]+\sum_1^n(\alpha_k-a_k)^2.$$

所得等式右边的第一项与数 α_k 无关．因此，当取 α_k 使第二项成为最小时，右边才成为最小，而这显然只有在数 α_k 分别等于傅里叶系数 a_k 时才能做到．

这样，我们得到了下列重要的结果．如果函数 φ_k 成为在线段 $[a, b]$ 上的一个规格化正交系，那么要和式 $\sum\limits_1^n \alpha_k\varphi_k(x)$ 在这线段上按照平均平方的意义最好地逼近函数 $f(x)$，必须而且只须数 α_k 是函数 f 关于 $\varphi_k(x)$ 的傅里叶系数．

根据等式(23)，容易证实函数

$$\frac{1}{\sqrt{2\pi}},\quad \frac{\cos x}{\sqrt{\pi}},\quad \frac{\sin x}{\sqrt{\pi}},\quad \frac{\cos 2x}{\sqrt{\pi}},\cdots$$

成为在线段 $[0, 2\pi]$ 上的一个规格化正交系．因此把上述结论应用到三角函数上得结果如下．

在 n 阶三角多项式

$$t_n(x)=\alpha_0+\sum_1^n(\alpha_k\cos kx+\beta_k\sin kx)$$

中，对已给的以 2π 为周期的连续函数所算出的傅里叶和 $S_n(x)$，在线段 $[0, 2\pi]$ 上按照平均平方的意义逼近函数 $f(x)$．

由所得结果以及 §7 中所讲到的费耶尔定理，我们还发现一件巧妙的事实．

设 $f(x)$ 是以 2π 为周期的连续函数，$\sigma_n(x)$ 是它的 n 阶费耶尔和[由 §7 中的等式(27)所决定]．

引用记号

$$\max|f(x)-\sigma_n(x)|=\eta_n.$$

因为傅里叶和 $S_k(x)$ ($k=0, 1, \cdots, n$) 是阶数为 $k\leqslant n$ 的三角多项式，所以显然 $\sigma_n(x)$ 是 n 阶三角多项式. 因此根据以上所证明的和 $S_n(x)$ 的最小性质，下列不等式成立:

$$\int_{-\pi}^{\pi}[f(x)-S_n(x)]^2dx$$
$$\leqslant\int_{-\pi}^{\pi}[f(x)-\sigma_n(x)]^2dx$$
$$\leqslant\int_{-\pi}^{\pi}\eta_n^2dx=2\pi\eta_n^2.$$

因为由费耶尔定理，数量 η_n 当 $n\to\infty$ 时趋近于零，所以我们得到下列重要结果.

对于以 2π 为周期的任一连续函数，下列等式成立.

$$\lim_{n\to\infty}\int_{-\pi}^{\pi}[f(x)-S_n(x)]^2dx=0.$$

在这种情形下，还可以说当 n 无限增大时，连续函数 $f(x)$ 的 n 阶费耶尔和在平均平方的意义下趋近于 $f(x)$.

可是对于更广泛的一类函数，即本身及其平方都是勒贝格可积的函数，实际上上述事实仍成立.

我们讲到此为止，并且将不举出傅里叶级数及正交函数论中根据平方逼近的方法所得到的另外一些有趣的结果. 在第六章中读者已经认识到正交函数系的一些重要的物理实例. 最后，我们指出，在第十九章(第三卷)所作的另外一些讨论中，在实质上也很注意到这个问题.

文　献

专　著

Н. И. 阿赫叶慈尔，逼近论讲义，科学出版社，1957 年版。

这本书除了包含函数逼近论中的基本结果以外，还包含着一系列更特殊、更精密的问题。

В. Л. 冈察洛夫，函数插补与逼近理论，科学出版社，1958 年版。

И. П. 纳唐松，函数构造论，科学出版社，1958 年版。
以上两书包含函数逼近论的基础。

С. М. Никольский, Приближение многочленами функций действительного переменного, Математика в СССР за тридцать лет (1917—1947), Гостехиздат, 1948.
本文详细评述苏联数学家在这方面的工作。

余家荣 译
秦元勋 校

第十三章　近似方法与计算技术

§1. 近似及数值的方法

近似方法的特点　在许多情况中能运用数学来研究外界现象，是因为控制这些现象的规律具有量的性质，并且能写成某些公式、方程或不等式的形式．这就可能用数值方法研究现象本身并做出对实践所需要的计算．

求到量的规律以后，就可以纯粹用数学的方法来研究．为了叙述确切起见我们将考察写作方程形式的某一规律．例如牛顿力学中物体运动的规律，热传导或电磁振荡的规律等等．这些方程已在第五、六章中详细讨论过．对于过程通常还要附加某种形式的条件(第五、六章中这是边界及初始条件)，由这些条件可以决定方程个别的解．

这里首要且基本的数学问题如下：

(1)　问题可解性的确立．即使问题的可解性从物理观点上看来是相当明显的，但严格表述出来的问题的可解性的数学证明就是问题的数学提法本身的正确性的证据．在广泛类别中可解性是能确立的．

(2)　对于描述被研究现象特征的量，试求它用公式写出的显性表达式．这种表达式通常只能在少数最简单的情况中找到．并且常常碰到所得解的显性表达式是如此复杂，要用它们来研究现象及求特性的必要的数据就非常困难，有时甚至完全不可能．

(3)　指出可用以建立近似公式的步骤，该公式会给出具有任何给定准确度的解．这一点在广泛类别的情况中能做到．

(4)　最经常的是可能指出一个或几个方法，用来数值地求问题所需的解．

自然科学与技术问题解的数值方法以及在一定程度上近似方法的发展为数学中一个特殊分支的任务，这一分支现在通常称为计算数学.

计算数学的方法当然是近似的方法，因为每一个量我们只能利用计算求到一定有效数字位数的准确度，例如到十进位的五、六位等等.

实用上这已足够，因为知道某一量的精确值往往并非必要.例如技术问题中的未知量通常用来决定制造中主体产品的大小及别的参量.每一产品只是近似地制造出来的，因此技术上的计算精确度超过产品所能达到的"许可"程度显然是没有意义的.

因此没有必要按绝对精确的公式来进行计算，而探求未知量时也没有必要解绝对精确的方法.精确的公式及方程全然容许以不精确的来代替，只要它们是如此地接近精确的，足以相信由这种代换所产生的差错不会超出规定的界限.

以后我们将重新讨论这里所述的，一些问题用另一些代替的可能性.现在我们想再强调一下计算方法的第一个特点——它们按自己的特性通常只能给出近似的结果而在实践上也只需要这种结果.

我们再注意到数学中数值法的另一特点.在任何计算中只能对有限个数进行运算，而计算后只得到有限多个结果.因此每一个准备用数值法来解的问题，应该预先化成这样的形式，就是要经过有限多次的算术运算能够得到全部的结果.假如我们依某一公式来进行计算，那么它应该预先变换，使得其中只保留含有有限多个参数的有限多项，例如，众所周知，很多函数能被表作幂级数的和

$$f(x)=c_0+c_1x+c_2x^2+\cdots. \qquad (1)$$

如函数 $\sin x$ 就能展开为幂级数

$$\sin x=\frac{x}{1!}-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots,$$

其中 x 是角的弧度单位的度量.

为了求 $f(x)$ 的精确值，我们应该求级数(1)"全部"项的和，一般说来这是不可能的．要近似地求 $f(x)$，只需取级数的某些有限个项就够了．例如，可以证明，对从零到半个直角之间的角度计算 $\sin x$ 准确到 10^{-5} 只要在幂级数中取到 x^5 项，即以多项式 $\frac{x}{1!}-\frac{x^3}{3!}+\frac{x^5}{5!}$ 近似地代替 $\sin x$．

在定某一未知函数的数学分析问题的数值求解中应该以某种方式以关于求几个数值参数的另一问题来代替本来的问题，这些参数知道了就可以近似地计算未知函数．以例说明之．

设在区间 $a\leqslant x\leqslant b$ 上要解对于微分方程

$$L(y)-f(x)=y''+p(x)y'+q(x)y-f(x)=0 \qquad (2)$$

在边界条件 $y(a)=0,\ y(b)=0$ 下的边值问题．一种可能的解法，即伽辽尔金方法是从满足边界条件(第六章 § 5)的某组线性无关的函数系 $\omega_1(x),\ \omega_2(x),\ \cdots$ 出发的．选择在这种意义下的"完全"系，即在 $[a,\ b]$ 上的可积函数中只有各点(更确切地说，"几乎处处")都等于零的函数才能与所有 $\omega_k(k=1,\ 2,\ \cdots)$ 正交．$y(x)$ 满足微分方程(2)这一条件可以写成正交性条件的形式

$$\int_a^b [L(y)-f]\omega_k\,dx=0 \quad (k=1,\ 2,\ \cdots). \qquad (3)$$

设问题的解可以展开为 ω_k 的级数

$$y(x)=a_1\omega_1(x)+a_2\omega_2(x)+\cdots, \qquad (4)$$

这儿应当求出系数 a_k．对任何 a_k 级数(4)的和总满足边界条件．留下的是要选择 a_k 使得适合等式(3)．系数 a_k 的全体是无穷集，要全部计算出来，一般地说，是不可能的．为了简化起见只保留(4)式右部的有限多项，并近似地置

$$y(x)\approx a_1\omega_1(x)+\cdots+a_n\omega_n(x). \qquad (5)$$

因为我们只有 n 个任意参数 a_k，所以不能使它对所有 $\omega_k(k=1,\ 2,\ \cdots)$ 满足等式(3)．因此我们被迫抛弃微分方程(2)的精确解．但应该期待，假如 n 取得足够大而条件(3)对开首 n 个函数 ω_k 适合时和(5)将以微小的误差来满足这个微分方程．这就导出伽辽

尔金方法中所列的方程

$$\int_a^b \left[L\left(\sum_{k=1}^{n} a_k \omega_k \right) - f \right] \omega_i dx = 0 \quad (i=1,\ 2,\ \cdots,\ n).$$

从这些方程求出 a_k，我们就作出函数(5)的近似表达式.

在变分问题依李茨方法求解中，在函数的近似调和分析以及许多别的问题中公式都可以类似地化简.

我们再提出把方程简化的一个例子. 假设需要通过解某一函数方程(例如微分或积分方程)来求单变数或多变数函数 y. 可以选取某组点(网格上的)上的值 $y_1,\ y_2,\ \cdots,\ y_n$ 作为决定函数 y 的参量.

我们应该用关于 n 个未知量 $y_k(k=1,\ 2,\ \cdots,\ n)$ 的数值方程组来代替函数方程. 这种替代通常可以用许多方法来实现. 并且永远应该关心数值方程组的解与函数方程的解差别要足够微小.

下面是这种替代的一些例子. 在用欧拉方法解一阶微分方程 $y'=f(x,\ y)$ 时，我们以递推的数值方案来代替这个方程，这方案可用前一值近似地求得未知函数的下一值(第五章 § 5)，

$$y_{n+1}=y_n+(x_{n+1}-x_n)f(x_n,\ y_n).$$

用网格法求拉普拉斯方程 $\Delta u=\dfrac{\partial^2 u}{\partial x^2}+\dfrac{\partial^2 u}{\partial y^2}=0$ 的近似解时，这个方程用线性代数方程组(第六章 § 5)

$$u(x+h,\ y)+u(x,\ y+h)+u(x-h,\ y)+u(x,\ y-h) \\ -4u(x,\ y)=0.$$

来代替.

再观察一个这类的例子. 设要数值地解积分方程

$$y(x)=f(x)+\int_a^b K(x,\ s)y(s)ds. \tag{6}$$

我们要在那里求未知函数 $y(x)$ 值的一些点记作 $x_1, x_2, \cdots, x_n$. 对于作出用来代替(6)的数值方程组时，并不需要使得这方程对区间 $a\leqslant x\leqslant b$ 中所有的点都满足，而只要在点 $x_i(i=1,\ \cdots,\ n)$ 上有

$$y(x_i)=f(x_i)+\int_a^b K(x_i,\ s)y(s)ds.$$

然后用某一以 $x_1, \cdots, x_n$ 为分点的近似积分和(按辛卜森公式或某一别的公式[1])来代替积分

$$\int_a^b K(x_i, s)y(s)ds \approx \sum_{j=1}^{n} A_{ij}K(x_i, x_j)y(x_j).$$

我们得到用来决定未知值 $y(x_i)$ 的线性代数方程组

$$y(x_i)=f(x_i)+\sum_{j=1}^{n} A_{ij}K(x_i, x_j)y(x_j) \qquad (i=1, 2, \cdots, n). \quad (7)$$

可以看出,所有上述方法都用探求某些参数来代替探求函数,这些参数近似地决定函数. 因此这些方法的准确度与函数由这组参数所决定的好坏程度有关,例如函数能被(7)型的表达式逼近的好坏程度或函数用它在某组点上的值来表示的好坏程度有关. 这类问题联系到数学的一个特殊分支,叫做函数逼近论(见第十二章). 由此可见函数逼近论对应用数学具有巨大的意义.

近似方法的收敛性及误差的估计 我们将更详细地来讨论从计算观点上提出的对于近似方法的要求. 最简单而根本的要求是以选定的准确度来求未知量的可能性.

按问题性质的不同,计算的精确度可以有很大的不同. 对某些粗糙的技术上的计算往往精确到十进位有效数字二、三位就够了. 大部分工程上的计算完成到十进位数三、四位. 科学上的计算往往需要大得多的精确度. 一般说来,精确度的要求与时代的前进同时增长.

因此,那种可以求得任意精确结果的近似方法及步骤对于计算具有特殊意义. 这些方法通称为**收敛的**. 因为它们在计算的实践中最常碰到,而对它们提出的要求也最典型,所以今后就来考察它们.

设 x 为未知量的精确值,在每一个这种方法中可以作出逼近于解 x 的数列 $x_1, x_2, \cdots, x_n, \cdots$.

一旦作逼近的方法指出之后,这个方法理论方面的首要问题就是建立逼近于解的收敛性 $x_n \to x$,而假使收敛性并非永远成立,

1) 见第十二章, § 3.

那么阐明一些条件，在此条件之下收敛性成立.

当收敛性确立之后，就产生了更难、更深刻的收敛速度的估计问题，也就是当 $n\to\infty$ 时估计 x_n 多快地趋向于解 x. 每一个收敛的方法在原则上可以求得任意高精确度的解，只要求到对应于足够大 n 的近似值 x_n. 但照例 n 愈大，消耗在寻找 x_n 上的劳动量将更大. 因此，假如 x_n 缓慢的趋近于 x，那么为了达到必需的精确度可能要耗费过多的计算时间.

在数学本身，特别是在其应用中，大家很熟悉许多情况，就是虽然当求解 x 时能够指出一个收敛的过程，但是它需要这样大的，甚至有了近代快速计算机[1]也不可能完成的计算量.

不够快的收敛性是通常判定方法缺点的标志之一. 但是这种标志当然不是唯一的，当比较各种方法的时候应该注意到问题的许多方面，特别是在机器上进行计算的方便性. 从两个方法中有时宁愿挑选收敛较慢的方法，如果按照这个方法的计算在计算机上容易实现.

当 x 用近似值 x_n 代替时得到的误差等于差 $x-x_n$. 它的精确值并不知道，为了要判定收敛的速度就从上来估计差的绝对值，就是作出量 A_n 使满足

$$|x-x_n|\leqslant A_n,$$

并称之为它的误差估计. 下面我们将引出估计 A_n 的例子. 通常依据估值 A_n 随 n 增长而递减的速度来判断 x_n 收敛于 x 的速度.

---

1) 这儿是一些计算过程收敛很慢的简单例子. 大家知道，级数 $\frac{1}{1}-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots$ 收敛，而其和等于 2 的自然对数. 利用这级数能够近似地求 $\ln 2$，只要计算足够的前 n 次的和 $s_n=\frac{1}{1}-\frac{1}{2}+\cdots\pm\frac{1}{n}$ 就行. 但是能够算出，对于计算 $\ln 2$ 误差小于第五位有效数字的一半时，应该取级数的十万项以上. 求这么多项数的和，假如只利用台式计算机就非常困难. 级数

$$\frac{1}{\sqrt{2}}=1-\frac{1}{2\cdot 1!}+\frac{1\cdot 3}{2^2\cdot 2!}-\frac{1\cdot 3\cdot 5}{2^3\cdot 3!}+\frac{1\cdot 3\cdot 5\cdot 7}{2^4\cdot 4!}-\cdots$$

也是大家所熟悉的. 它收敛得这样慢，利用它计算 $\frac{1}{\sqrt{2}}$ 准确到 10^{-6} 大致就要取 10^{10} 项，这甚至在快速计算机上也是费力的.

假如估值要反映出 x_n 对 x 的真实接近度，就必须要 A_n 与 $|x-x_n|$ 相差甚微．此外估值 A_n 应该是有效的，也就是它要事实上能被求到，否则它是不能应用的．

设 x 为数值变量而其值要从某一方程来决定．假定我们已把方程变作形状

$$x=\varphi(x). \tag{8}$$

我们用**迭代法**来解这个方程．它也常称作**逐次逼近法**，为了解释方法本身以及和它相关的误差估计，我们将讨论一个数值方程的情况．（这个方法也能用到数值方程组，用到微分与积分方程以及许多其他情况中．读者已在第五章 § 5 中碰到这个方法在常微分方程上的运用．）

我们假定用某种方法已求得方程根的近似值．假如 x_0 是方程(8)的精确解，那么把它代入方程的右部 $\varphi(x)$ 之后我们应该得到结果，它等于 x_0．但一般说来 x_0 不等于精确解，因此代入的结果将与 x_0 不同．把它记作 $x_1=\varphi(x_0)$．

为了确立在什么情况下 x_1 就比 x_0 更接近于精确解，我们转向问题的几何涵义．

研究函数

$$y=\varphi(x). \tag{9}$$

取数轴并假定用这轴上的点表示 x 与 y．等式(9)使每一点 x 对应同一轴的某一点 y．这也可以看作规定数轴自身点变换的规律．

在数轴上给出区间 $[x_1, x_2]$．在变换(9)之下点 x_1 及 x_2 变为点

$$y_1=\varphi(x_1) \text{ 及 } y_2=\varphi(x_2).$$

区间 $[x_1, x_2]$ 变为区间 $[y_1, y_2]$．比值

$$k=\frac{|y_2-y_1|}{|x_2-x_1|}$$

就是在变换(9)下的区间“伸长系数”．若 $k<1$，就产生区间的压缩．

回到方程(8)．它说明所求的点 x 经变换(9)后应该变为自身．因此解方程式(8)等价于在数轴上求这样一些点，它们在变换(9)

之下变为自身，也就是保持不动.

现取区间$[x, x_0]$，它的一端放在不动点x上，而另一端放在点x_0上. 在变换下x_0变到x_1而区间$[x, x_0]$变到区间$[x, x_1]$. 假如φ是这样的函数，使在变换(9)下任何区间要产生压缩，点x_1就一定比x_0更接近于方程(8)的根.

为要得到收敛于(8)的精确解的逼近值，只要重复地代入(8)的右部就作出一列数

$$x_1=\varphi(x_0),\ x_2=\varphi(x_1),\ \cdots,\ x_{n+1}=\varphi(x_n),\ \cdots. \tag{10}$$

下面我们证明逼近列(10)的收敛性定理[1].

假设函数$\varphi(x)$在区间$[a, b]$上给定而等式(9)给出$[a, b]$的自身变换，也就是对每一个属于$[a, b]$的点x，$y=\varphi(x)$仍旧属于$[a, b]$. 我们仍在$[a, b]$上取初始近似x_0；所有逼近列(10)就也在区间$[a, b]$上. 在这些条件下下面的定理要成立. 若$\varphi(x)$具有导数φ'，在$[a, b]$上满足条件

$$|\varphi'|\leqslant q<1,$$

则下面的推断成立. 方程(8)在$[a, b]$上有根x^*. 数列(10)收敛于此根，而收敛性的速度由估值不等式

$$|x^*-x_n|\leqslant\frac{m}{1-q}q^n$$

表征出来，其中$m=|x_0-\varphi(x_0)|=|x_0-x_1|$. 方程(8)在$[a, b]$上有唯一的根.

为了要证明这个断言我们来估计差x_2-x_1. 假如用泰勒公式[第二章(第一卷)，§ 9，(26)]，其中令$n=0$，得到

$$x_2-x_1=\varphi(x_1)-\varphi(x_0)=\varphi'(\xi_0)(x_1-x_0).$$

点ξ_0在x_1与x_0之间，显然是属于区间$[a, b]$的. 因此$|\varphi'(\xi_0)|\leqslant q$而

$$|x_2-x_1|\leqslant q|x_1-x_0|=mq,$$

同理

1) 这个以及别的和它类似的定理由于上述几何意义在数学中常称为压缩映射定理.

$$|x_3-x_2|=|\varphi(x_2)-\varphi(x_1)|=|\varphi'(\xi_1)(x_2-x_1)|$$
$$\leqslant q|x_2-x_1|\leqslant mq^2.$$
继续这种估值,就找到在每个值 n 要满足的不等式
$$|x_{n+1}-x_n|\leqslant mq^n. \tag{11}$$
现在来确立数列 x_n 的收敛性. 为此研究辅助级数
$$x_0+(x_1-x_0)+(x_2-x_1)+\cdots+(x_n-x_{n-1})+\cdots. \tag{12}$$
它的头 $n+1$ 项的部分和等于
$$s_{n+1}=x_0+(x_1-x_0)+\cdots+(x_n-x_{n-1})=x_n.$$
因此 $\lim\limits_{n\to\infty}s_{n+1}=\lim\limits_{n\to\infty}x_n$ 而 x_n 具有有限极限就等价于级数(12)的收敛性. 现将级数(12)与级数
$$|x_0|+m+mq+\cdots+mq^{n-1}+\cdots$$
相比较. 由于估值(11), 级数(12)各项的绝对值不大于上述级数的对应项. 但是这个级数,假如除开第一项$|x_0|$, 便是以 q 为公比的几何级数,又因 $q<1$,故级数收敛. 级数(12)因而也是收敛的,而数列(10)将收敛于某有限极限 x^*,
$$\lim_{n\to\infty}x_n=x^*.$$

显然 x^* 属于区间$[a, b]$,因为所有 x_n 都是属于此区间.

若在等式$x_{n+1}=\varphi(x_n)$中命$n\to\infty$则趋于极限即得到等式$x^*=\varphi(x^*)$, 这说明了 x^* 实际上满足方程(8). 现在来估计 x_n 对 x^* 的接近程度. 取 x_n 及任何以下的逼近 x_{n+p}
$$|x_{n+p}-x_n|=|(x_{n+p}-x_{n+p-1})+(x_{n+p-1}-x_{n+p-2})$$
$$+\cdots+(x_{n+1}-x_n)|\leqslant mq^{n+p-1}+mq^{n+p-2}$$
$$+\cdots+mq^n=\frac{mq^n-mq^{p+n}}{1-q}.$$
由此,当 $p\to\infty$,因为 $x_{n+p}\to x^*$ 及 $q^{n+p}\to 0$, 推得
$$|x^*-x_n|\leqslant\frac{m}{1-q}q^n.$$

下面验证唯一性. 设 x' 为方程在$[a, b]$上的另一解. 估计差 $x'-x^*$ 的值.
$$|x'-x^*|=|\varphi(x')-\varphi(x^*)|=|\varphi'(\xi)(x'-x^*)|\leqslant q|x'-x^*|,$$

因此 $(1-q)|x'-x^*|\leqslant 0$.

因为 $1-q>0$, 上述不等式只有当 $|x'-x^*|=0$ 时才有可能. 就是 x' 与 x^* 一致.

证明了的定理不仅指出了使迭代法收敛的充分条件, 同时也提供一种可能性来估计计算所需的步数, 也就是假如精确解 x^* 用 x_n 来代替, 为了得到所需的准确度必须取多大的 n. 这种估值是有效的, 因为实际上借助于对于函数 φ 的研究可以求出不等式 $|x^*-x_n|\leqslant\frac{m}{1-q}q^n$ 中的量 m 及 q.

现举许多应用中常见的方程 $x=k\operatorname{tg}x$ 为例.

为了确定起见研究 $k=0.5$ 的情况. 设要求方程 $x=\frac{1}{2}\operatorname{tg}x$ 最小的正根. 只要利用函数 $\operatorname{tg}x$ 的任何表或图就容易断定它在点 1 的附近而稍稍比 1 大一些. 为了保证迭代法收敛性定理中条件 $|\varphi'|\leqslant q<1$ 的成立, 把函数 $\operatorname{tg}x$ 逆转而研究与原方程等价的方程 $x=\operatorname{arctg}2x$.

列出计算的结果. 取值 $x_0=1$ 为初始近似. 以下的近似值用 $\operatorname{arctg}x$ 函数表计算出来, 它们的数值被求出如下:

$$
\begin{aligned}
x_1 &= \operatorname{arctg} 2 &&= 1.10715\\
x_2 &= \operatorname{arctg} 2.21430 &&= 1.14660\\
x_3 &= \operatorname{arctg} 2.29320 &&= 1.15959\\
x_4 &= \operatorname{arctg} 2.31918 &&= 1.16370\\
x_5 &= \operatorname{arctg} 2.32740 &&= 1.16498\\
x_6 &= \operatorname{arctg} 2.32996 &&= 1.16538\\
x_7 &= \operatorname{arctg} 2.33076 &&= 1.16550\\
x_8 &= \operatorname{arctg} 2.33100 &&= 1.16554\\
x_9 &= \operatorname{arctg} 2.33108 &&= 1.16555\\
x_{10} &= \operatorname{arctg} 2.33110 &&= 1.16556\\
x_{11} &= \operatorname{arctg} 2.33112 &&= 1.16556
\end{aligned}
$$

在这儿计算就停止, 因为往下迭代只会重新出现根

$$x^* = 1.16556.$$

根的逼近几何图形如图1所示．这儿 x_n 趋于 x^* 是这样的快，在图上 x_4 已知 x^* 相会合了．

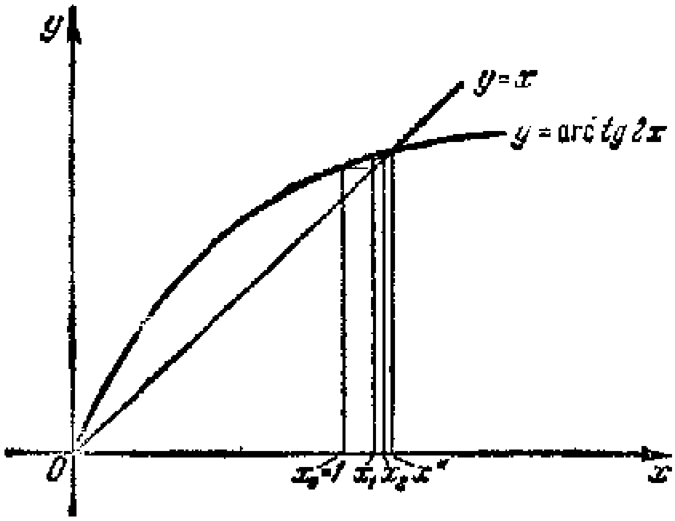


图 1

再举出一个用迭代法的例子．数值地求解积分方程

$$y(x) = e^x - \frac{1}{6}\frac{1}{x+1}(e^{x+1}-1) + \frac{1}{6}\int_0^1 e^{xt}y(t)dt, \qquad (13)$$

它的精确解是 $y=e^x$．

首先用线性代数方程组来代替积分方程．为此，用点 $t=0$, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1 把积分区间分为四等分．未知函数 y 在这些点上的值相应地用 y_0, y_1, y_2, y_3, y_4 来表示．假如要求方程在 $x_0=0$, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1 上被满足，而积分对四个部分区间用辛卜森求积和来代替[第十二章, § 3, 公式(6)]，那么得到下列方程组

$$y_0 = \frac{1}{6}[0.083333\,y_0 + 0.333333\,y_1 + 0.166667\,y_2 + 0.333333\,y_3 + 0.083333y_4] + 0.713619,$$

$$y_1 = \frac{1}{6}[0.083333\,y_0 + 0.354831\,y_1 + 0.188858\,y_2 + 0.402077\,y_3 + 0.107002\,y_4] + 0.951980,$$

$$y_2=\frac{1}{6}[0.083333\,y_0+0.377716\,y_1+0.214004\,y_2+0.484997\,y_3$$
$$+0.137393\,y_4]+1.261867,$$
$$y_3=\frac{1}{6}[0.083333\,y_0+0.402077\,y_1+0.242499\,y_2+0.585018\,y_3$$
$$+0.176417\,y_4]+1.664181,$$
$$y_4=\frac{1}{6}[0.083333\,y_0+0.428008\,y_1+0.274787\,y_2+0.705667\,y_3$$
$$+0.226523\,y_4]+2.185861.$$

方程组用迭代法来解．取相应方程的自由项作为 $y_k(k=0, 1, 2, 3, 4)$ 的初始近似值：$y_0^{(0)}=0.713619$, $y_1^{(0)}=0.951980$, … 以后各次近似中求得的值列在表 1 中．

表 1

近似的次数	y_0	y_1	y_2	y_3	y_4
1	0.93423	1.20841	1.56129	2.01542	2.59972
2	0.98517	1.26699	1.62905	2.09419	2.69173
3	0.99667	1.28021	1.64433	2.11194	2.71245
4	0.99926	1.28319	1.64778	2.11595	2.71713
5	0.99985	1.28386	1.64856	2.11685	2.71818
6	0.99998	1.28402	1.64873	2.11705	2.71842
7	1.00001	1.28405	1.64877	2.11710	2.71847
精确解的值	1.00000	1.28403	1.64872	2.11700	2.71828

表 1 的下端列出解的精确值用以比较．进一步的近似将不能改善 y_k 已求得的值．y_k 后面几个数位的偏差是被用求积和来代替积分所产生误差的影响所决定的．

方法的稳定性　计算对于近似方法的理论还提出了一个普遍的要求，由于它有重大的意义不得不加以说明．这就是计算过程稳定性的要求．问题的实质如下．每一个近似方法引导到某一计算方案．常常发生这样的情况，为了获得所有需要的数必须依据这个方案进行一列冗长的计算步骤．每一步计算并非绝对精确地被完成，而只算到一定位的有效数字，因此每一步就形成某一微小的误差．所有这些误差将影响到以后的结果．

选取的计算方案可能显得如此不恰当，在开始计算中所容许的微小差错，于计算推进的过程中在结果上将显出愈来愈强的影响，而在较远的部分引起了对精确值很大的偏差。

设讨论微分方程

$$y'=f(x, y)$$

在初始条件 $y(x_0)=y_0$ 下的数值求解，设需求等距点 $x_k=x_0+kh$ $(k=0, 1, \cdots)$ 上 $y(x)$ 的值。

设计算已开始进行到第 n 步，并作表于下。现在我们求 y_{n+1}。在欧拉折线法中近似地置

$$y_{n+1}=y_n+hy'_n. \tag{14}$$

x	y	$y'=f$
x_0	y_0	y'_0
x_1	y_1	y'_1
$\cdots$	$\cdots$	$\cdots$
x_{n-1}	y_{n-1}	y'_{n-1}
x_n	y_n	y'_n

这儿只根据列于上表末行的数 y_n 及 y'_n 就可以求出 y_{n+1}。我们为了要提高求 y_{n+1} 时的精确度利用表上最后两行已知的数。那么可以作出计算公式

$$y_{n+1}=-4y_n+5y_{n-1}+h(4y'_n+2y'_{n-1}). \tag{15}$$

我们注意，假使绝对精确地进行计算，也就是说带有无限多个有效数字，那么每当函数 y 为线性多项式时公式(14)就给出正确的结果；公式(15)对每个三次及三次以下的多项式也是正确的。初看起来似乎用公式(15)所得的结果应该比用折线法求得的结果要精确些。然而容易看出公式(15)并不适合于计算，因为它的运用将引起误差很快的增长。

导数值 y'_n 与所在 y'_{n-1} 项含有微小的因子 h，故这些值的误差比起 y_n 及 y_{n-1} 中的将显出较小的影响。为了简化起见，假定 y' 精确地求得，而在观察一般误差的过程中不考虑它们。假设在求

y_{n-1} 时我们出了差错 $+\varepsilon$，在求 y_n 时出了差错 $-\varepsilon$. 那么正如等式(15)所示在 y_{n+1} 中形成了差错量 $+9\varepsilon$. 在 y_{n+2} 中差错将为 -41ε 而以后将飞快的增加. 公式(15)导致对误差为不稳定的计算过程，故该被抛弃.

下面所列的例子足以信服地说明，计算方案的不稳定性对结果能导致如何强烈的歪曲. 微分方程 $y'=y$ 在初始条件 $y_0=1$ 下已被解出. 精确解为 $y=e^x$. 对于数值求解自变量 x 被取作以步长 $h=0.01$ 的等距值，也就是 $x_k=0.01k$. 近似解用两个方法计算出来：按折线法(14)及按公式(15). 为了好作比较在表 2 中给出到七位十进数字的精确解.

表 2

x	精确解的值	算出的近似解的值	
		按公式(14)	按公式(15)
0.00	1.0000000	1.0000000	1.0000000
0.01	1.0100502	1.0100000	1.0100502
0.02	1.0202013	1.0201000	1.0202012
0.03	1.0304545	1.0303010	1.0304553
0.04	1.0408108	1.0406040	1.0408070
0.05	1.0512711	1.0510100	1.0512899
0.06	1.0618365	1.0615201	1.0617431
0.07	1.0725082	1.0721353	1.0729726
0.08	1.0832871	1.0828567	1.0809789
0.09	1.0941743	1.0936853	1.1056460
0.10	1.1051709	1.1046222	1.0481559
0.11	1.1162781	1.1156684	1.3996456
0.12	1.1274969	1.1268250	−0.2808540

按公式(15)所求得解的近似值在开始几步要比由折线法求得的结果精确些. 但是公式(15)的不稳定性甚至经过不多几步就大大歪曲了近似值 y_k 并且导致与 y_k 真值相差很大的数值.

近似方法的选择 每个计算归根到底都能化为四则运算——加、减、乘及除来完成. 指出计算方法，这就意味着为了要得到所期望的结果，应该取哪些初始值以及哪些算术运算按照哪样的次

序来执行．我们希望在很简单的计算实例子中说明，计算的组织工作与负责计算准备的数学家的经验与知识的关系是多么大，以及假如运用特殊的、适当地选出的计算方法将能够得到怎样的结果．

假设应该解带 n 个未知数 $x_1, x_2, \cdots, x_n$ 的 n 方程组

$$\begin{aligned}
&a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1,\\
&a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n=b_n.
\end{aligned}$$

在代数方程组的理论中(第十六章§3)由行列式给出了未知量数值的显性表达式

$$x_j=\frac{\Delta_j}{\Delta}\quad (j=1,\ 2,\ \cdots,\ n). \tag{16}$$

这里 Δ 为方程组的行列式，

$$\Delta=\begin{vmatrix}a_{11}\ a_{12}\cdots a_{1n}\\ a_{21}\ a_{22}\cdots a_{2n}\\ \cdots\cdots\cdots\cdots\\ a_{n1}\ a_{n2}\cdots a_{nn}\end{vmatrix},$$

而行列式 Δ_j 为由 Δ 的第 j 列以方程组的自由项代换所得．

假定我们想用公式(16)来解方程组，并且在行列式通常的定义基础上来计算而不用任何简化的方法．为此我们应该做多少次乘与除的运算呢？(因为加法与减法比较简单我们将不加注意)．我们要计算 $n+1$ 个 n 阶行列式．其中每一个行列式都有 $n!$ 项，并且每一项为 n 个因子的积，它的计算需要 $n-1$ 个乘法．关于计算所有的行列式我们就需完成 $(n+1)n!(n-1)$ 个乘法．必需的乘与除法的总数就等于 $(n^2-1)n!+n$.

现在我们选择别的方法来解方程组，就是用消去法．这个方法的计算方案是由高斯提出的．从方程组的第一个方程求出 x_1

$$x_1=\frac{a_1}{a_{11}}-\frac{a_{12}}{a_{11}}x_2-\cdots-\frac{a_{1n}}{a_{11}}x_n.$$

为此需要 n 个除法．把 x_1 代入其下 $n-1$ 个方程中的每一个都需要做 n 个乘法．消去 x_1 并作出带未知数 $x_2, \cdots, x_n$ 的 $n-1$ 方程组总共需要 n^2 个乘法及除法．继续这些计算，求到用消去法来算出全部值 $x_j(j=1, \cdots, n)$ 所需要的乘法与除法为 $\frac{n}{6}(2n^2+9n-5)$ 次．比较这两个结果．对解五个方程的方程组在第一种情况中需要 2885 次乘法及除法，而在第二只要 75 次．

对十个方程的组的运算次数将相应地为 $(10^2-1)10!+10\approx 360,000,000$ 及 $\frac{10}{6}(2\cdot10^2+9\cdot10-5)=475$．可见计算工作的分量可能强烈的依赖于计算方法的选择．当筹划计算工作的时候合理地选择方法往往能够大大缩减工作量．

§2. 最简单的计算辅助工具[1]

表 最老的计算辅助工具为数学表．最简单的表，例如乘法表以及对数与三角函数值的表，当然读者们都是很熟悉的．在实际的事业中需要求解的问题的范围是不断地扩张的．新问题常常应用新的公式来解，或者会引致新的函数，因此必需的表的个数经常在增加．

每一个表不管它的构造如何总含有早先做出的结果，它作为一种特殊的数学记忆．印刷的或手抄的表是专供人来查阅使用的．但也可以提到特殊形式构成的表，例如在特殊的穿孔卡上的表．它们适宜于计算机工作上的应用，但碰到这种表的机会少得多，所以我们不预备讨论它．

最普遍的是函数数值表．假如函数 y 与一个变量 x 有关，那么对应于它的最简单的表具有形式

1) 本节描述最简单的计算工具及机器．近代快速计算机的叙述留待第十四章．由于篇幅的限制我们也没有涉及图解法．

x	y
x_1	y_1
x_2	y_2
…	…
x_n	y_n

(17)

这所谓一个进口的表[1]. 从表中能够直接取出所列 x 值的对应值. 未列出的 x 所对应的值利用各类插值公式求得, 关于这些公式在第十二章已提到[2]. 常常在表上适当地附以除函数值之外的一些辅助量的值, 这些量使插值容易完成. 通常这是一次或二次差分的值. 更特殊的表需要特殊地用慎重选定的插值公式, 并在表中列入适当的数据.

当二元函数 $u=f(x, y)$ 列表的时候它的值通常编为两个进口 x 及 y 的表中, 其图形式如

$x \backslash y$	y_1	y_2	…	y_m
x_1	u_{11}	u_{12}	…	u_{1m}
x_2	u_{21}	u_{22}	…	u_{2m}
…	…	…	…	…
x_n	u_{n1}	u_{n2}	…	u_{nm}

(17′)

这种表的每一列就是一个进口表, 而表(17′)是许多形如(17)的表的结合. 照例二元函数表的分量往往比具有相同独立变量步长的单元函数表的容量要大得多. 由于这点二元函数的列表比起单元函数来就少得多.

下面简单的计算说明表的篇幅当变量个数增加时可能增长多快. 设要列四元函数 $f(x, y, z, t)$ 的表, 并且每一个独立变量总

1) 表所占的列可能很长. 在印刷中为了安置它, 把它分成一系小段, 并把小段用某种方式编组. 然而在这种情况下仍旧称表为一个进口的.

2) 照例, 表上值 x_i 愈少插值就愈复杂, x_i 取得相隔愈近就愈容易. 对插值速度的要求可能是很不相同的. 供射击用的表非常希望插值"一眼"几乎瞬刻就能完成. 在提高了精确度供科学计算用的表中就容许需要一系列运算为插值.

共取 100 个值. 设函数需要用较低的准确度来计算——只要三位有效数字. 假若在这种条件下单元函数被列成表, 整个数值表由一百个三位数组成, 就很容易被安置在一页纸上.

但在四个进口的数表中对于函数 f 将与 x, y, z, t 值的组合 100^4 有同样多的 f 值. 由此容易估计出来表要占 300 多卷书这么大的篇幅.

由于表非常冗繁的缘故, 多元函数只在甚少而特别简单的情况下才被列表. 在最近几年才开始系统地研究几类多元函数, 对于它们能够编出进口数小于独立变量个数的表. 同时开始了研究一些规律使这类表的结构能够尽可能地简单一些.

举出这种函数的最简单的例子.

设需列出 x, y, z 三元函数 u 的表, 该函数具下列结构

$$u=f[\varphi(x, y), z].$$

显然, 假如引进辅助变量 $t=\varphi(x, y)$ 并把 u 表作复合函数

$$u=f(t, z),$$

$$t=\varphi(x, y),$$

则用两个具有两个进口的数表就够了.

为了便于运用这些表, 它们可以用下面的方法结合起来. 研究函数 $t=\varphi(x, y)$ 并关于 y 解此方程

$$y=\Phi(x, t),$$

原则上说来, 对于函数 $t=\varphi(x, y)$ 或 $y=\Phi(x, t)$ 中的哪一个列表都是一样的, 但就其中第二个函数列表对我们更方便些. 对函数 $y=\Phi(x, t)$ 及 $u=f(t, z)$ 作两个具有两个进口数表, 并把它们结合为如下的表

x_1	x_2	$\cdots$	x_i	$\cdots$	t	z_1	z_2	$\cdots$	z_k	$\cdots$
					t_1					
					t_2					
					$\vdots$					
$\cdots$	$\cdots$	$\cdots$	y_j		t_j	$\cdots$	$\cdots$	$\cdots$	u_{jk}	$\cdots$

求对应于给定值 x_i, y_j, z_k 的 u 值可进行如下：找出记录着数 x_i 的一列，并沿着它移动找到值 y_j(或接近于它的值)，沿水平线对着它的就是对应的 t 值．假如沿这个水平线更向前移动，那么在 z_k 列我们找到需要的值 $u=f(x_i, y_j, z_k)$．

在这个例子里，利用两个具有两个进口的表以及简单的运用规律就足以代替一个具有三个进口的表．

利用各种把表缩短的可能性使得在某些情况下减少表的篇幅比起进口数等于独立变数个数的表来，表的篇幅要减少到几十、几百甚至几千倍．

台式计算机 几乎和数学表同样古老的计算工具为各种计算设备．它们在古代希腊就被应用了．在苏联早就出现了算盘，并且现在仍广泛地应用着．

最初的一些计算机模型是在十七世纪由巴斯卡、莫尔兰德及莱布尼茨所做的，从那时起机器被多次改变及改进，在上世纪末叶，尤其在本世纪初它们得到大大的推广．

我们只讨论在某些类型的机器上，并尝试说明它们所提供的加速计算的可能性．我们从小的或台式通用计算机开始．无论构造如何它们中间的每一种都专为执行完成四则运算，并且乘和除法是用多次按位的加及减法来完成的．

这种机器初期典型的模型是大家很熟悉的具有奥得纳轮的奥得纳欧算术计算机(图 2)．在安置机构上数的排置是由每一个数位上的排数杆运动到所需数的刻度上来完成的．做加法时被加数排在安置机构上，然后把摇臂转一圈就输送到结果计数器上，在此它自动的加到那儿已有的数上去．减法由摇臂在反方向旋转来进行．乘法由被乘数排在安置机构上，然后在乘数的每一位上把它多次相加来完成．例如，乘上 45 就相当于被乘数五次相加，然后把被乘数移进一位再四次相加．

对于数的除法被除数应该放在结果计数器上，然后从它把除数多次按位相减就求得了商．结果由摇柄在每位的回转次数来决定．并且能够在计算机的回转次数计量器上读出．

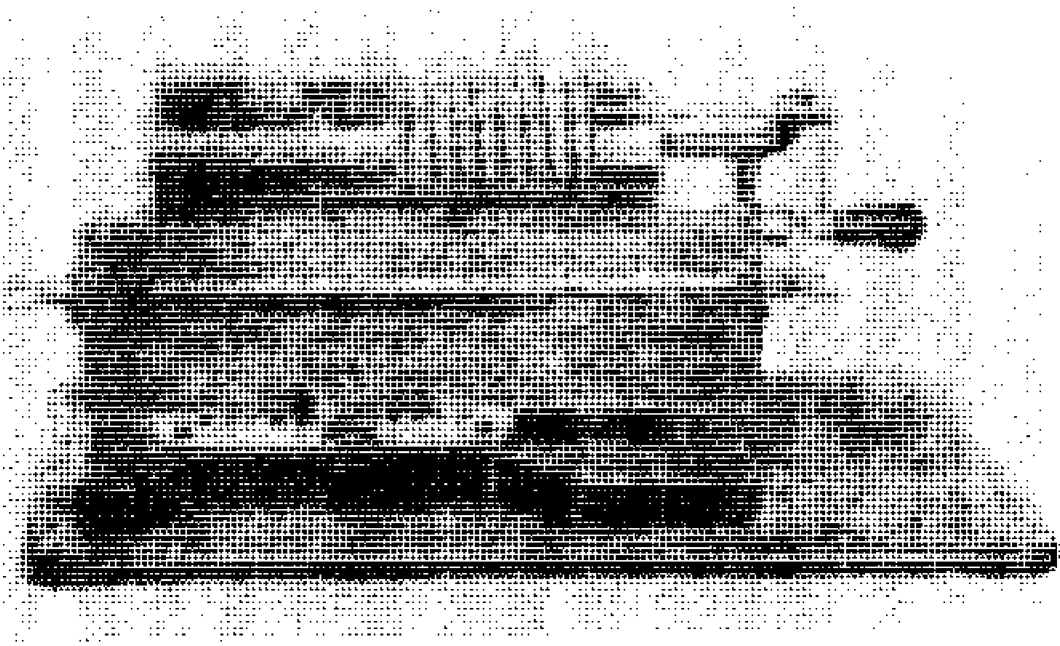


图 2

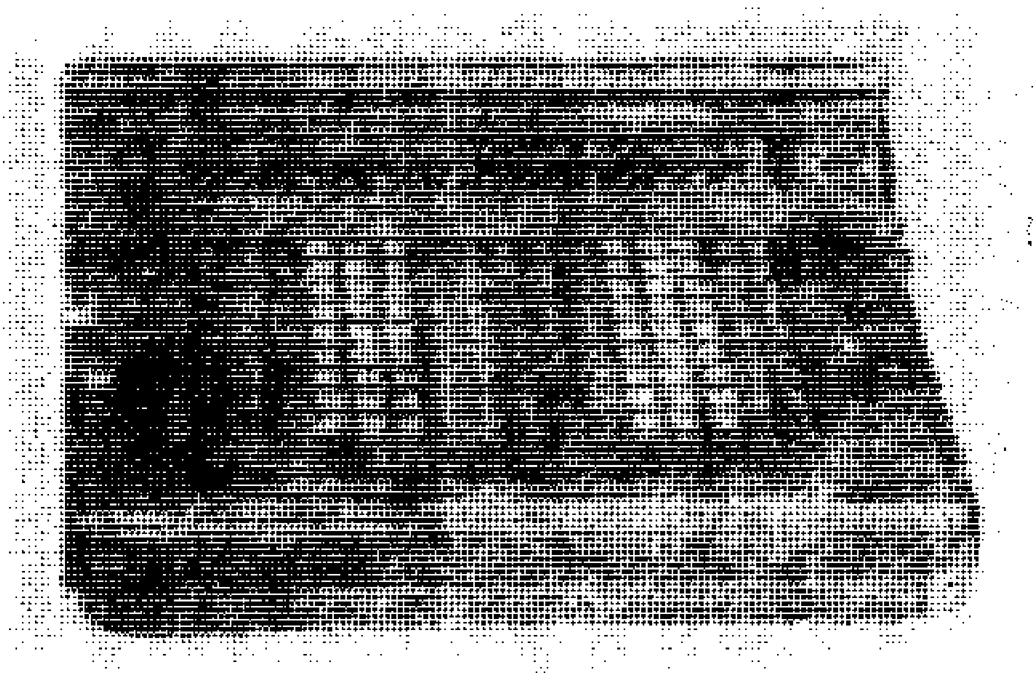
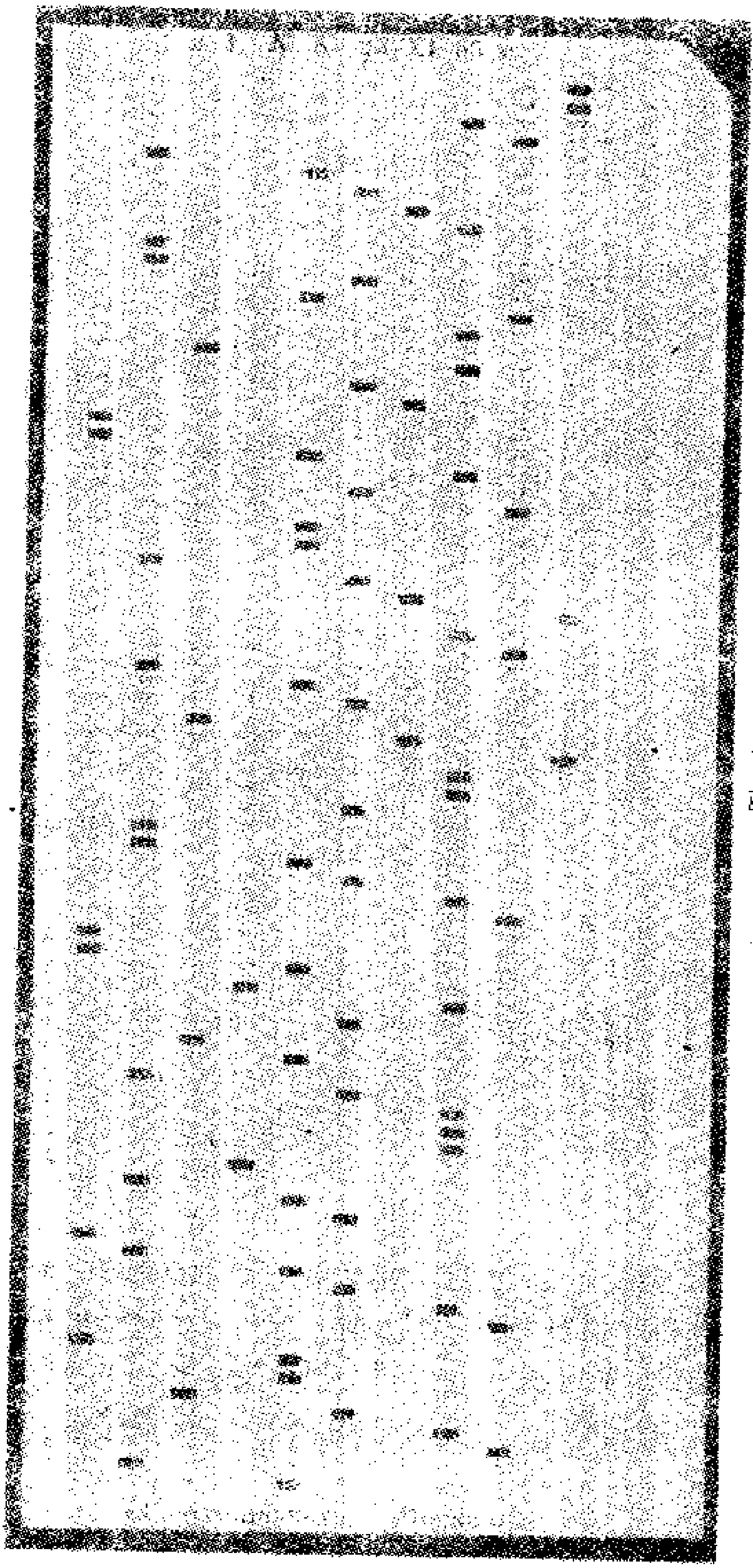


图 3

我们简单地回忆了在算术计算机上的计算顺序，只是为了给出以后改进台式机的更明显的方向．某一些改进的目的在于使得机器更方便些，而不改变它的结构原则上的设计．这类的改善例如电动的算术计算机，可以加速机器的工作并解除了计算员转动摇柄的劳动．

为了加速及简化数在安置机构上的排置引进了触键结构．数



字的安置不是由杆子回转一定刻度数来完成的，而只要单单压一压相应的键就行．已经造出这样的自动计算机，在它上面工作的时候计算员只要安置被运算的数字然后压一下按钮指示要完成四种运算中的哪一种(图 3)．留下的事情不用人来过问，机器自动做完．台式计算机的改善也保证了它们工作速度的大大提高，以致于在最完善的机构上乘法的结果只要在开动机器后一秒钟就可得到．机器工作的进一步加快显然是多余的，因为计算员自己执行安置数字及记录结果的操作需要比较长的时间．

分析计算机及继电式机器 分析计算机是造出来为统计计算以及为了财政与工业核算上用的．它们是专为进行大量同一类型不复杂的计算用的．对于执行技术及科学上的计算它们就不太方便，这是因为运算的“记忆”容量非常小以及对它们作出计算的程序的可能性是有限制的．虽然有这些缺点，直到快速电子机出现以前分析计算机在复杂及大量的计算中仍被相当广泛地应用，只要当计算能化为不大的一系列能被大量完成的运算．

分析计算机的数被记录在穿孔卡片上(图 4)．利用在确定位置上打孔的办法在卡片上引入数码及符号．卡片被送进机器里通过一列电刷子．在电刷子下面若经过卡片上打穿了的孔，它就使电路闭合并使机器的这一部件或那一部件开始工作．

机器总是成套地工作，并且每套至少由下列机器组成．

穿孔器——作为在卡片上打孔之用．机器有手按的键盘并以打字机的速度进行工作．

分类器——用来把卡片按次序分堆，它们该按那种次序被引进计算机中去．工作的速度是每分钟 450—650 张卡片．

再生穿孔器或再生器——把打穿的孔从一些卡片上移到另一些上，比较两堆卡片并依一定的标帜在其中进行选择．工作的速度大约每分钟 100 张卡片．

列表器(图 5)——执行加法及减法运算，此外还印出计算结果．每小时可以加工 6—9 千张卡片．

乘积穿孔器(倍积器)——把数相加、相减及相乘．结果以卡

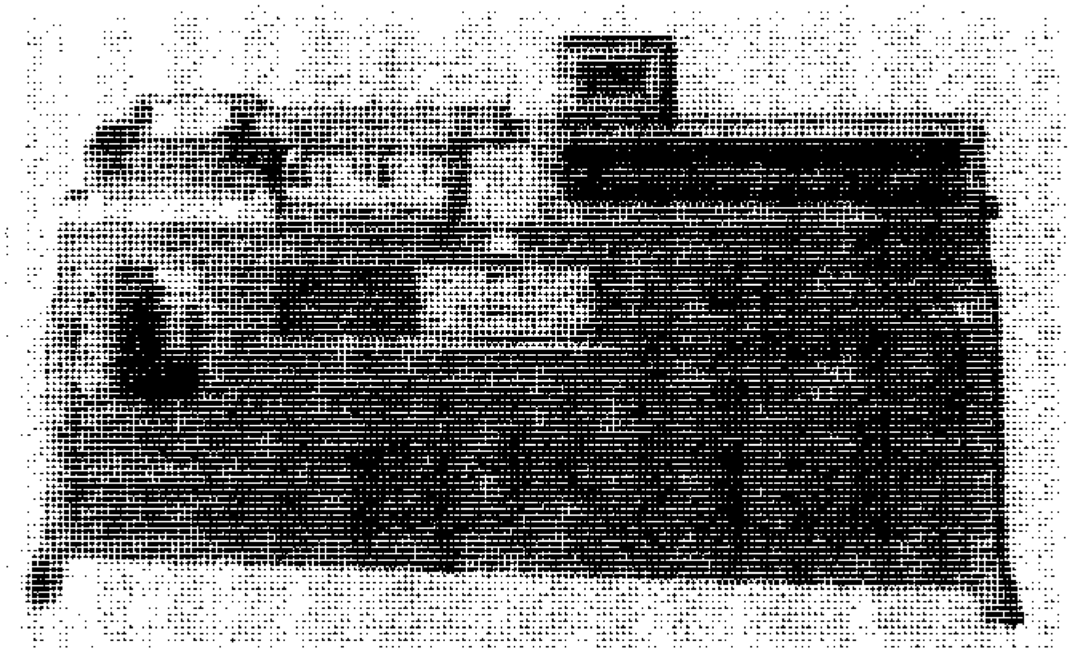


图 5

片上穿孔的形式给出来．在对6—7位的数进行工作时每小时可以完成700—1,000个乘法．

分析计算机工作起来是比较慢的．假如粗略地估计它们能够完成的工作量，那么可以说上列一套分析计算机能够代替12—18个台式自动计算机．试制更快速的计算机器的最初努力的结果就是，继电式机器，它是基于运用电力机械的继电器而做成的．这种机器的工作速度大约比分析计算机的工作速度要大十倍．但是更大的优点却在别的方面：继电式机器能执行复杂的计算程序并且具有非常灵活的控制．这使得大大推广了可以在机器上解的技术与科学问题的范围．但是几乎在继电机器出现的同时就创造了第一批用程序来控制的电子计算机，在这种机器上它使工作速度得到进一步的跃进．转移到电子技术上来时增加速度的可能性有多大，可以依据下列数字来判断：电子管状态改变所需的时间是用百万分之一秒作为单位来测量的．

连续式数学机器 连续式数学机是一种这样构造的物理系统(机械构造、电路等等)，使得在系统的连续改变的参数(位移、回转角、电流、电压等等)之间能实现那样的数值关系，正如所需求解的数学问题中的量之间的数值关系一样．这类机器常称为**模拟机**．

每个连续式的机器是专用来解决某一小类问题的.

机器给出的解的精确度与零件制造的质量、机器的装配、调整、工作中的惯性误差等等有关. 根据长期运用机器的经验断定这种机器通常能够给出解中 2—3 位正确的有效数字. 在这一方面模拟机就大大的赶不上数字计算机, 后者在原则上具有无限的计算精确度.

连续式机器的重要性质为它们便于解决大量的同一类型的问题. 此外, 它们常常能够比数字计算机以快得多的速度给出解答来. 它们主要的优越性是在许多场合中更便于引入问题的初始数据, 并且得到的结果表为更方便的形式.

存在很多种形式的模拟机. 许多问题可以模拟, 并且它们的每一个都可能用几种方法来模拟, 例如利用机械或电的结构以及别的办法. 可以创造一些机器或机器的部件来模拟个别的数学运算, 例如加法、乘法、积分法、微分法等等. 各种计算公式也可以模拟, 例如, 制造计算多项式值的以及函数在调和分析中的傅氏系数的机器. 同样可以制造模型, 用来复制数值的或函数的方程. 从完全不同领域的问题之间多种相似之处导致同样的微分方程. 方程的同一性使得, 例如, 可能用电来模拟热的现象而热工问题就用电的测量的办法来解, 毫无疑问这要有利得多, 因为电的测量比热的精确并且实行起来要容易得多.

由于模拟机数量很多要想用几行字不仅不可能来描述机器本身, 甚至也不可能描述它们的结构原理. 为了使读者对如何模拟数学问题具有初步概念, 我们将提供两个简单数学机器的概要的描述, 其中的一个用来求函数的积分, 另一个则用来求拉普拉斯方程的近似解之用.

阻力积分仪(图 6) 正如它本身的名称所示, 是用来求函数的积分用的. 它靠磨擦力的作用来工作. 结构原理的示意图如图 7, 其中 1——积分仪的机座, 2——水平地安置着的带有轴的阻力盘, 3——阻力小轮, 即小轮有压平的边缘并且不仅能够沿着盘子滚动, 而且在垂直于滚动平面的平面里移动. 零件 4 与 5 组成了螺旋

机构，机件上的螺丝帽 4 与携带小轮的支架相连接着．假如螺旋的螺距记作 h，则当螺旋回转 γ 角度时小轮在图形的平面上移动距离 $\rho=h\gamma$.

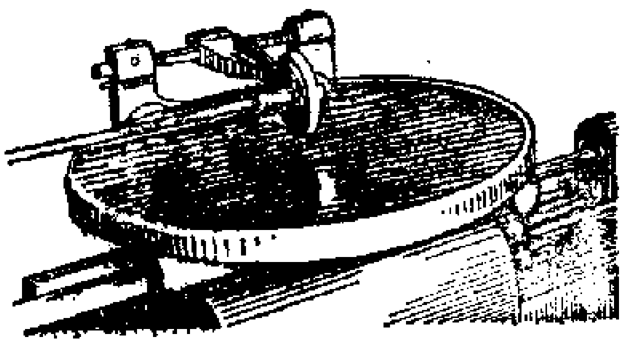


图 6

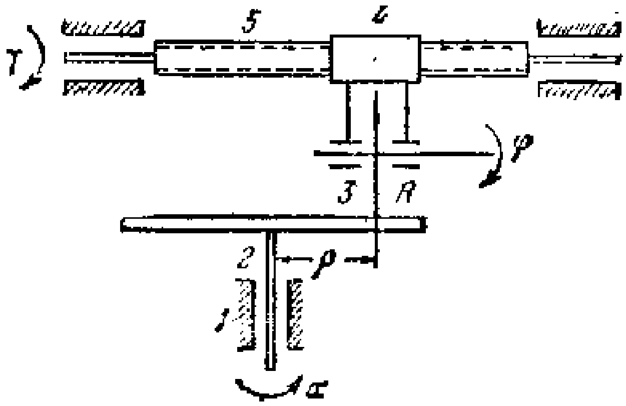


图 7

设盘的轴转 $d\alpha$ 角度．此时小轮子的接触点移动了 $\rho d\alpha$ 弧长．假如小轮子无滑动的沿着盘子滚动，则小轮的回转角就等于

$$d\varphi=\frac{\rho}{R}\,d\alpha=\frac{h}{R}\,\gamma\,d\alpha,$$

假设盘子的轴从角度 α_0 开始旋转并且小轮的初始回转角是 φ_0．由上式积分之求得

$$\varphi-\varphi_0=\frac{h}{R}\int_{\alpha_0}^{\alpha}\gamma d\alpha.$$

取定角度 γ 及 α 之间适当的关系之后，我们在广泛类别的情况中可以用阻力积分仪来计算积分．利用积分机器许多微分方程就可能用机械方法求解．

我们举出第二个例子．设在平面上给定区域 Ω，它由曲线 l 所限定．需要求函数 u，它在区域内满足拉普拉斯方程

$$\Delta u=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0,$$

而在回路 l 上取给定的值

$$u|_l=f.$$

作正方网格点

$$x_k=x_0+kh,\ y_k=y_0+kh,\ k=0,\ \pm1,\ \pm2,\ \cdots,$$

而区域 Ω 本身用由正方形组成的多边形来代替．相应地回路 l 用折线来代替．把边界值 f 及 l 转移到折线上去．未知函数 u 在结点 $(x_j,\ y_k)$ 上的值表作 u_{jk}．在 Ω 中求拉普拉斯方程近似解通常以

在区域的所有内点上应该满足的代数方程组来代替它:

$$u_{jk}=\frac{1}{4}[u_{j+1,k}+u_{j,k+1}+u_{j-1,k}+u_{j,k-1}].$$

为了求解此方程组可以作下列电网模型．在平面上作二向度导电网络，电路图如图8．结点间的电阻认为是一样的．假设在网络区域的边界结点加上数值等于这些结点上边界值的电压．它们在网络的所有内点上也引起了电压．在结点(x_j, y_k)上的电压记作$U_{j,k}$．若在结点(x_j, y_k)运用克希霍夫定律,则显然在这结点上应该满足方程

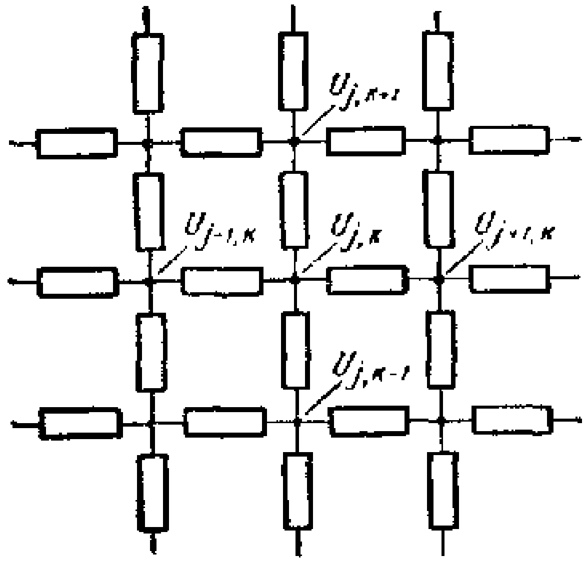


图 8

$$\frac{1}{R}[(U_{j+1,k}-U_{j,k})+(U_{j,k+1}-U_{j,k})+(U_{j-1,k}-U_{j,k})$$
$$+(U_{j,k-1}-U_{j,k})]=0,$$

它与上面所示对应的方程组只是写法形式上不同而已．代数方程组解的值在网格的结点上应该与电压U_{jk}相一致，并且能够从此模型上用普通电学测量的办法求得．

文　　献

Л. В. Канторович и В. И. Крылов, Приближенные методы высшего анализа. Изд. 4-е, Гостехиздат, 1952.

书中叙述了偏微分方程近代数值解法．

Л. Коллатц, Численные методы решения дифференциальных уравнений, М., 1953.

А. Н. Крылов, Лекции о приближенных вычислениях, Изд. 6-е, Гостехиздат, 1954.

曾在近似方法的进一步发展中起了很大作用的第一本教科书．

Э. Уиттекер и Г. Робинсон, Математическая обработка результатов наблюдений, Перев. с англ., Изд. 2-е, ОНТИ, 1935.

本书包含丰富的历史上的资料．

Хаусхольдер, Основы численного анализа, ИЛ, 1956.

通 俗 书 籍

Б. Н. Делоне, Краткий курс математических машин, Гостехиздат, 1952.

Ф. А. Виллерс, Математические инструменты, ИЛ. 1949.

描述各种连续式的数学仪器.

Г. П. Евстигнеев и И. С. Евдокимов, Счетно-цифровые машины, М., 1953.

叙述分析计算机.

Статьи в Большой советской энциклопедии: Графические методы, Номография. Таблицы.

М. С. Тукачинский, Как считают машины, Гостехиздат, 1952.

储钟武 译

冯 康 校

第十四章　电子计算机

§1. 电子计算机的功用和基本工作原理

数学方法广泛地应用在科学与技术中，可是许多重要问题的解都与巨量的计算有关，因此当利用普通手摇计算机进行计算时，这些问题实际上是不可能得到解决的。以空前的速度进行计算的电子计算机的出现，在应用数学解决物理学、力学、天文学、化学等最重要问题方面实现了一个变革。

现代的通用电子计算机在一秒钟内能完成成千上万次算术和逻辑运算，它可以代替几万和几十万计算人员的劳动。利用这样的速度来计算，例如，炮弹的弹道计算速度就比炮弹本身飞行的速度要快得多。

除了能以高速度进行算术及逻辑运算外，同一台通用计算机还能够解决各式各样的问题。这种计算机在本质上是一种新的工具，它不仅能大大地提高劳动生产率，而且也能解决过去被认为是不可能解决的问题。

在许多场合下必须快速进行计算，才能使所取得的结果具有实际的价值。这点在对次日的可靠的天气预报的例子中是显而易见的。用人工计算一昼夜的可靠天气预报时需要几天几夜的时间。显然在这种计算速度下所得的结果便失去了实际的价值。应用电子计算机做天气预报的计算便能完全并及时地解决这个问题。

快速电子计算机　在苏联科学院精密机械及计算技术研究所中所试制出的快速电子计算机(БЭСМ)，就是这种机器的一个例子(图 1)。此计算机在一秒钟内平均能完成八千至一万次算术运算。不能不提醒一下，一位有经验的计算员利用手摇计算机在每一班内(八小时)只能完成二千次算术运算。这样一来，一台计算机在

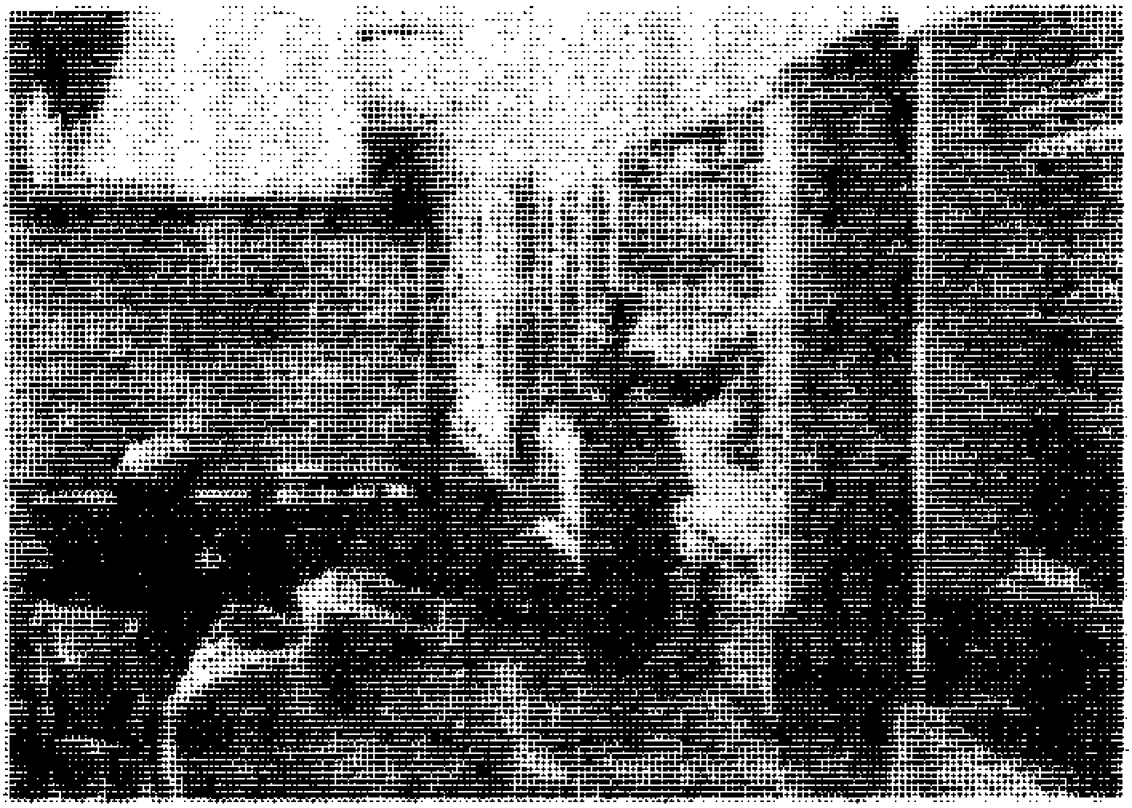


图1 БЭСМ 电子计算机的全貌

几小时内能完成的计算是一位有经验的计算员一辈子也做不完的．这台机器能代替由几万人组成的计算大军，光是为了容纳这些人，就得需要有几十万平方公尺的房屋．

在这台电子计算机的运行期间已经解决了科学与技术的不同领域的大量问题，由此为国家节省了几亿卢布．我们举几个例子．

在国际天文历方面，几天内计算出了太阳系内约七百个小行星的运行轨道，并且考虑到了木星与土星对它们的影响，确定了它们在十年内的坐标，准确地算出了它们每隔四十天所在的位置．从前这种计算需要一个大计算局几个月的工作．

在根据地区测地图数据来制地图时需要解决具有大量未知数的代数方程组．具有 800 个方程的问题大约需要进行二亿五千万次算术运算，这个问题在电子计算机上不到二十小时便解决了．

在计算机上计算出了用来决定运河最陡的而又不塌散的侧面的各种形式的换算表，因而能使得在水力工程建设中节省大量物质和时间．从前有 15 个计算员一起工作，试图解出这个问题的一

个方案，但经过几个月的光景也没有成功．在电子计算机上十个方案的计算花了不到三小时．

在机器上可以反复试验这种或那种问题的解的多种方案，并从中选取最合适的一种．这样便能够决定桥梁的最有利的力学结构，找出机翼、喷气发动机、涡轮机叶片的最佳形式等等．

实际上计算的无限准确性能够使我们在电子计算机上算出各种各样的科学和技术用的表．为了使得在 БЭСМ 上计算出具有五万个弗兰涅尔积分值的表，只用一小时．

电子计算机在解逻辑问题方面的运用 电子计算机除了解数学问题外，也能解逻辑问题，例如利用计算机可以把文献从一种语言翻译成另一种语言．在这种情况下，机器中所存储的将不是数字而是代替辞典的单词和个别的词汇．

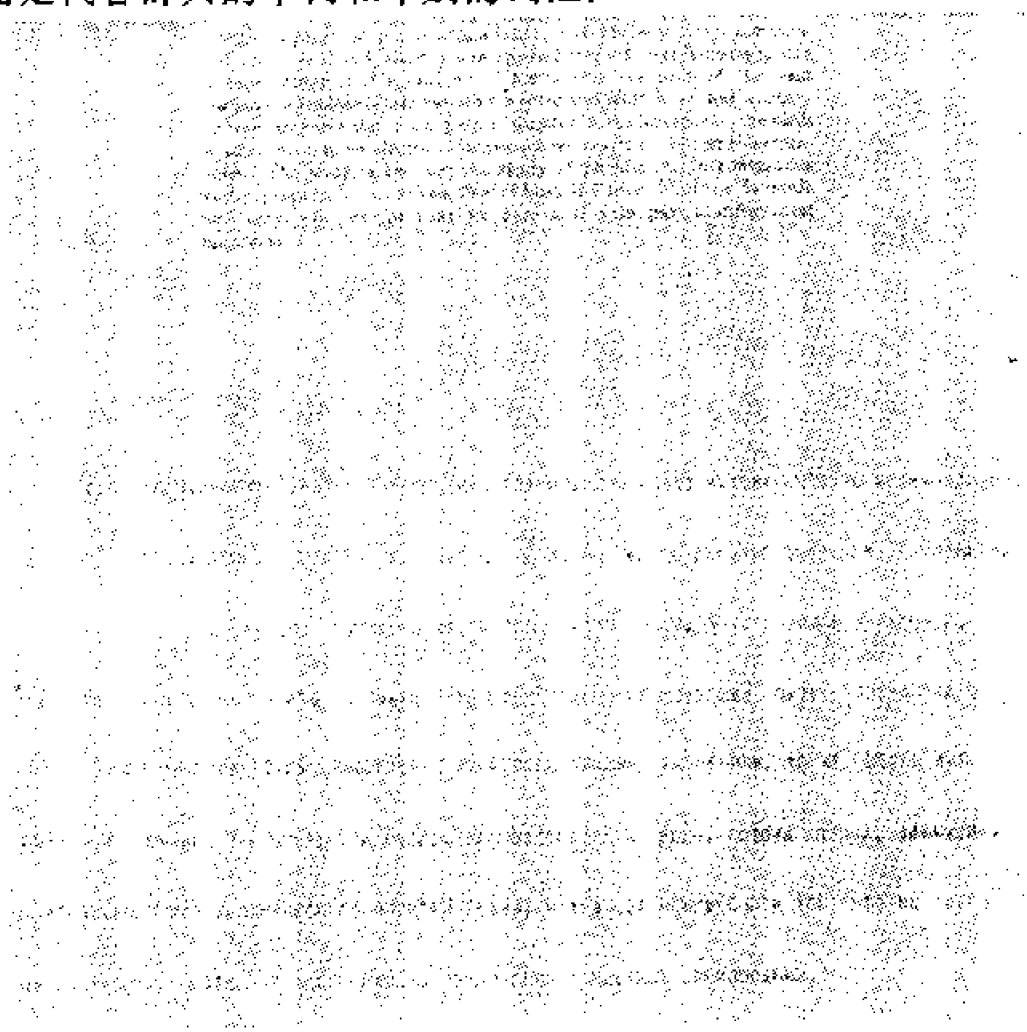


图2 翻译的例子

计算机通过原文中的单词与“辞典”中的单词相比较找出所需要的单词．然后计算机利用以程序形式记录下来的语法和句法规则对所找到的单词进行“加工”，按格、数、时等变化单词，并将它安放在句子中的适当地位，最后将译文印在纸上．为了达到成功的翻译，要求语言学家与数学家在编制程序方面付出大量和细致的劳动．

在苏联科学院已经编制了用来把科学技术文献从英文译成俄文的试验性字典和程序，并于1955年底在БЭСМ(此计算机并不是专门为了翻译而设计的)上做了第一次翻译试验(图2)．

做为试验，在БЭСМ上成功地解决了复杂的逻辑问题，例如象棋问题．对象棋游戏的全面分析在现代电子计算机上是做不到的，因为游戏中有巨量的可能组合．当采用近似法时，棋子可以用点数来估值，例如皇帝——一万点，皇后——一百点，堡垒——五十点等等．此外，对个别的局势因素，例如开线、重卒等也以适当的点数来估值．计算机经过一连串的测试选取那样一个方案，它在一定的步数内对于对方的任意的回棋给出最佳的点数关系．但是，由于有大量的可能组合，解不得不局限于比较少的步数的测试中，因此不能考虑到棋赛的战略部署．

电子计算机的基本工作原理　现代的电子计算机是电子自动装置的一个复杂的组合，在机器内运用了电子管，锗晶体元件、阴极射线管、磁性元件、光电管、电阻、电容器以及其它无线电技术零件．

算术运算是用电子计数线路以非常快的速度来执行的，这些电子计数线路组成为运算器(图3)．

为了保证高速计算光凭对数字进行快速的算术运算是不够的．因此在计算机内整个计算过程完全自动化，选取所需数字以及制定对数字做运算的一定的次序也都自动地完成．

需要做运算的数以及计算的中间结果都应该存储在机器内．用来做存储数字的装置——“存储器”——能够选取任一个所需要的数，以及接收计算结果．存储器的容量，即它所能存储数字的数量

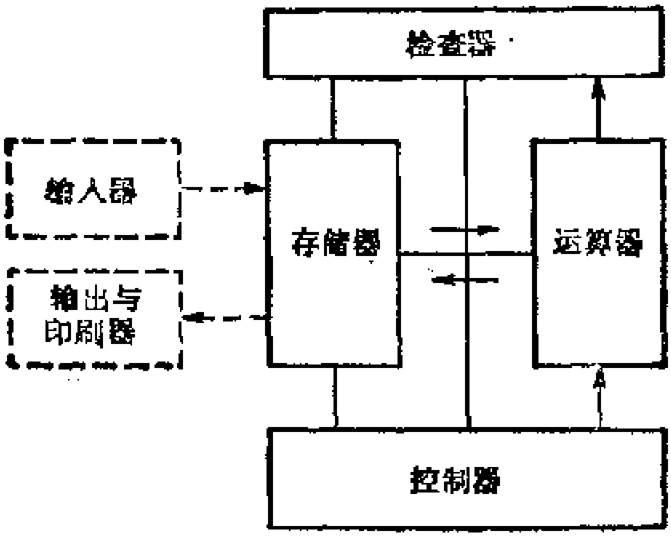


图 3　电子计算机基本部件的概括图

在很大的程度上决定着机器对解各种各样问题的应用的灵活性.

在现代的电子计算机内存储器的容量为 1,000—4,000 个数.

在电子计算机内控制器执行下列操作：从存储器选取所需的数,对这些数需要进行的运算,向存储器传送结果和向下一步操作转移等. 在计算程序和原始数据被送入计算机之后，控制器便可保证计算过程的完全自动化.

为了往机器内输入原始数据和计算程序以及为了将结果印在纸上,在计算机内设有专门的输入器和输出器.

在机器上进行计算时必须对所取得结果的正确性有信心，也就是必须检查所进行过的计算. 计算正确性的检查或由专门的检查器来实现，或通过适当的程序设计用逻辑和数学检查方法来实现,这种检查的最简单例子就是"双手计数",即两次计算并比较所得的结果.

在着手解这种或那种问题之前必须了解所研究过程的物理实质,并将问题表为代数公式、微分或积分方程或其它的数学关系的形式. 在应用仔细研究过的数值分析方法时，几乎什么时候都可以把这样的问题化为一连串的算术运算,这样一来,最复杂的问题也都可用算术四则来解决.

在手算时要执行随便哪一种算术运算时,必须取两个数,对它

们进行给定的算术运算，并写出所得的结果，这个结果或者用在进一步的运算中，或者就是所求的答案.

在电子计算机内所执行的也就是这些运算. 机器的存储器化分为一系列的存储单元，所有的单元都编有一个番号，为要选取某一个数时，必须给出“存储”这个数的单元的番号.

为要执行随便哪一个算术运算时，必须给出两个存储单元的番号，由此选取两个数，给出对这些数所需要进行的操作和一个应该接收所得结果的单元的番号. 用一定的代码形式来表示的上述任务便称为“指令”.

因此，问题的解就等于顺序执行一系列的指令. 这些指令组成一个计算程序，它通常也存储在上面提到的存储器里.

计算程序，即指令的集合，是事先由数学家们制定的，计算程序可保证在解题时所必须的算术运算的一定序列.

为求得问题的解许多问题要求几千万甚至于几亿次算术运算、因此在电子计算机内采用了这样一些方法，即是用相当少数的原始指令来完成大量的算术运算.

除了执行算术运算的指令外，在电子计算机内也规定有实现逻辑运算的指令，例如，利用比较两数的大小来从进一步计算的两种方案中选取一种.

程序的指令以及原始数据以约定的代码形式写出. 通常用穿孔的方法记录在穿孔卡片或穿孔带上，或以记录代码脉冲的形式记录在磁带上，然后把这些代码送进计算机并转至存储器，以后机器便自动地执行所给的计算程序.

计算结果也同样地记录下来，例如以代码脉冲的形式记录在磁带上. 特殊的译码印刷装置把记录在磁带上的代码变为数字并把它们印为表格形式.

在电子计算机上以非常快的速度执行复杂的计算工作能在脑力劳动的领域内提供节省，这种节省如同从前在体力劳动方面采用机器生产所提供的节省. 当然，根据先由人规定的程序进行工作的电子计算机本身就没有创造的可能性，因此应用机器的目的

不是用机器代替人.

电子计算机在科学研究机关、设计局、计划机构中的广泛应用对解决国民经济问题开辟了无限的可能性.

在计算机的工作原理、构造及其应用和操作的进一步发展的事业中在工程师和数学家的面前呈现着宽广的远景.

电子计算机在人的手中是一个有力的工具. 这些机器对于正在建设共产主义社会的国家的意义是难于估计的.

§2. 在快速电子计算机中的程序设计和代码的编制

程序设计的基本原理 **1.** 在电子计算机上进行计算时必须将问题近似解的所选取的数学方法表示为一连串的算术运算. 这些运算的实现在机器上是由计算程序来保证的, 前面已经说过, 计算程序是由一系列的指令组成的. 当然, 如果对每一个算术运算都规定一条指令, 则计算程序将非常庞大, 甚至于只是把它们写下来所需的时间就将几乎等于人工执行这些算术运算时所需的时间, 因此在程序设计时力求用比较少数的指令来保证大量算术运算的执行.

为了阐明指令的结构与编制程序的方法我们用一个最简单的例子来研究在人工解题时所需执行的那些操作.

我们详述一下用欧拉方法解带下列已知初始条件的一价微分方程的例子.

$$\frac{dy}{dx}=ay,\ y|_{x_0}=y_0. \tag{1}$$

根据这个方法, x 值的整个幅度被划分为等长的区间 $\Delta x=h$, 在每一个区间的极限内, 假定微商 $\frac{dy}{dx}$ 的值不变并且等于此区间起点的微商值[1]. 在此条件之下, 对于第 k 区间的运算按下列公式进行:

1) 实际上常微分方程的数值解是用更复杂的和更准确的公式来求的.

$$\left(\frac{dy}{dx}\right)_k = ay_k,$$

$$\Delta y_k = \left(\frac{dy}{dx}\right)_k h = (ah)\, y_k,$$

$$y_{k+1} = y_k + \Delta y_k,$$

$$x_{k+1} = x_k + h.$$

做完第 k 区间的计算后转至计算第 $k+1$ 区间. 计算从已知的初始值 x_0, y_0 开始. 运算的序列见表 1.

表 1 用欧拉方法解方程(1)所必需的运算

运算号码	待定值	公 式	计 算[1]
1	Δy_k	$(ah)\, y_k$	$(ah)\,(2)_{k-1}$
2	y_{k+1}	$y_k+\Delta y_k$	$(2)_{k-1}+(1)_k$
3	x_{k+1}	x_k+h	$(3)_{k-1}+h$
4	—	—	印出所得之值 x_{k+1}
5	—	—	印出所得之值 y_{k+1}
6	—	—	对 x 和 y 的新值重复运算, 从运算号码 1 开始.
7	—	—	当 x 达到值 x_n 时, 计算结束.

在人工计算时只用了前三个运算, 而其余的运算并不明写下来而是自然地默认的, 例如对下一区间开始重复计算的指令以及结束计算的指令等等. 在机器上计算时所有这些运算(运算 4—7)都要正确地规定出来. 这样一来, 在机器上除了执行算术运算外, 还应该规定有控制运算(运算 4—7). 控制运算或者具有完全确定的性质(如运算 4 和 5), 或者具有依赖于所得结果的条件性质(例如运算 6 和 7). 因为后两种运算彼此是联系着的(需要执行其中之一), 所以在机器上最后两个运算综合为一个(比较运算), 它的形成如下: "如果值 $x<x_n$, 则必须从运算 1 起重复计算. 如果值

1) 1 在"计算"栏里括号内的数字标明对于运算应选取哪一个运算的结果, 例如: 在第一运算(第一行)中, 需要用对上一区间 $(2)_{k-1}$ 执行第二运算(第二行)时所取得的值乘上值 ah; 在第二运算中需要将对前一区间 $(2)_{k-1}$ 进行运算所得的值加上对本区间 $(1)_k$ 进行第一运算所得的值.

在开始计算时在"给定值"栏内对运算 2 及 3 必须列出原始值 y_0 及 x_0.

$x \geqslant x_n$，则必须停止运算”．因此，进一步的计算序列将依赖于在计算过程中所得的现行值 x 的大小．

2．对表 1 内算术运算的研究证明：为要执行任意一个算术运算必须指出应该对数字进行哪种运算，应该取哪些数，应该把取得的结果送到哪里去，因为这个结果在进一步的计算中还会用得上．

数码保存在机器的存储器内，因此应该指出相应单元的番号：应该从哪里取数和往哪里送结果．这样便形成了最自然的“三地址指令系统”．

在三地址指令系统内代码的一定的位数用来标明运算，即给出运算，也就是指出对于数所要进行的运算(运算码)．指令码以外的数位分为三个相等的小组，称为“指令地址”(图 4)．第一地址码指出一个存储单元的番号，从此选取第一个数，第二地址码指出一个存储单元的番号，从此选取第二个数，而第三地址码指出应接收所得结果的存储单元的番号．

运算码	第一地址	第二地址	第三地址

图 4　指令的三地址系统结构

用来控制计算过程的指令码也可以用三地址系统来表示．例如，“送数付印”的指令在运算码中应该有表示这一操作的番号；在第一地址内——付印的数所在的存储单元的番号，在第三地址内——印刷装置的番号(在第二地址内不放代码)．决定计算过程的这样或那样进程的指令叫做“比较指令”．这个指令的运算码指出必须比较两个数，这两个数的番号在第一及第二地址内标出．如果第一数小于第二数，则必须转至执行其番号在比较指令第三地址中所指明的指令，如果第一个数大于或等于第二个数，则在完成本指令后必须转至执行按番号顺序的下一指令．

指令码和数码一样也都保存在机器的存储器里，并且，如果在没有关于变更计算进程的指令(例如在有比较运算)时，它们是按番号顺序一个跟着一个的．

我们考虑一下，方才分析过的例子的计算程序是怎样的．假定存储单元内的数码分配如下：

数值 ah 在第 11 单元内

数值 h 在第 12 单元内

数值 x_n 在第 13 单元内

数值 x 在第 14 单元内

数值 y 在第 15 单元内

操作单元是第 16 单元[1]．

根据上表我们得到下面的计算程序(表 2)．

表 2 用欧拉方法解方程(1)的计算程序

指令番号	指令码				附注
	运算码	第一地址	第二地址	第三地址	
1	乘法	11	15	16	$\Delta y_k=(ah)y_k$
2	加法	15	16	15	$y_{k+1}=y_k+\Delta y_k$
3	加法	14	12	14	$x_{k+1}=x_k+h$
4	印	14	—	1	将 x_{k+1} 印在第一部打印装置上
5	印	15	—	2	将 y_{k+1} 印在第二部打印装置上
6	比较	14	13	1	如果 $x<x_n$，则转至指令番号 1；如果 $x\geqslant x_n$，则转至下一指令，即指令番号 7
7	停止	—	—	—	计算结束

指令码保存在存储器内(在我们考虑的例子中为单元 1—7)．保存在存储器的第一单元内的指令进入机器的控制器，按照这个指令执行乘法，用第 15 存储单元的数乘上第 11 单元的数，即算出量 $\Delta y_k=(ah)y_k$．所得的结果送至第 16 操作单元．这个运算完成后，存储器的按番号顺序的下一个单元．即第二单元内的指令进入机器的控制器．按照这一指令求得量 $y_{k+1}=y_k+\Delta y_k$，它被送至第 15 单元，即是代替原来的值 y，同样按第 3 指令求得 x 的新值；第 4 和第 5 指令对于 x 及 y 的新得的值实现印刷；第 6 指令决定计

1) 放置在计算过程中所得的中间值的单元称为工作单元．

算过程的进程．按照这个指令将位于第 14 单元内的数与位于第 13 单元内的数进行比较，即将所得的值 x_{k+1} 与最终值 x_n 比较．如果 $x_{k+1}<x_n$，则对下一区间必须重复计算，即回到起始的指令上，在此情况下就是第一指令．当第一数小于第二数时所应该转到的指令号码在比较指令的第三地址中指出．如果计算已达到 $x_{k+1}\geqslant x_n$，则比较指令向按番号顺序的下一指令，即第 7 指令转移，它终止计算过程．

在计算开始以前应该往存储器引进指令码(送入单元 1—7)，常数码(送入单元 11—13)以及初始数据，即是初始值 x_0 和 y_0(送入单元 14 和 15 内)．

在完成第一区间的计算后在第 14 和 15 单元内已经由 x_1 及 y_1 代替了 x_0 及 y_0，它们就是下一区间的起点的变量值．这样一来，在重复程序的同样一些指令时就可完成下一区间的计算等等．

上述的例子可证明，由于循环地重复一系列的指令便能以比较短的程序完成大量的计算．程序个别段落的循环重复方法是在编制解题程序中广泛地运用的．

3．另一个常用的能大大缩短程序的方法是自动地变更某些指令的地址．我们用计算多项式的例子来解释这种方法的实质．

设需要计算出多项式

$$y=a_0x^6+a_1x^5+a_2x^4+a_3x^3+a_4x^2+a_5x+a_6$$

的值．在计算机上进行计算时将此多项式表为形式

$$y=(((((a_0x+a_1)x+a_2)x+a_3)x+a_4)x+a_5)x+a_6$$

是比较方便的．设系数值 a_0—a_6 分别放在存储单元 20—26 内，值 x 放在第 31 存储单元内．用显易的方法编制的计算程序如表 3.

我们清楚地看到，在程序内乘法运算与加法运算是交替进行的．除了第一指令外，其余的一切乘法指令都是一样的：即是用位于第 31 单元的数去乘位于第 27 单元内的数，而把结果送至第 27 单元．一切加法指令都具有相同的第一及第三地址．当由一个加法指令向下一加法指令转移时，第二地址内的单元番号加一：在第二指令内从第 21 单元取数，在第四指令内从第 22 单元取数等等．

表3 多项式的计算程序

指令番号	指令码				备注
	运算码	第一地址	第二地址	第三地址	
1	乘法	20	31	27	a_0x
2	加法	27	21	27	a_0x+a_1
3	乘法	27	31	27	$(a_0x+a_1)x$
4	加法	27	22	27	$(a_0x+a_1)x+a_2$
5	乘法	27	31	27	$((a_0x+a_1)x+a_2)x$
6	加法	27	23	27	$((a_0x+a_1)x+a_2)x+a_3$
7	乘法	27	31	27	$(((a_0x+a_1)x+a_2)x+a_3)x$
8	加法	27	24	27	$(((a_0x+a_1)x+a_2)x+a_3)x+a_4$
9	乘法	27	31	27	$((((a_0x+a_1)x+a_2)x+a_3)x+a_4)x$
10	加法	27	25	27	$((((a_0x+a_1)x+a_2)x+x_3)x+a_4)x+a_5$
11	乘法	27	31	27	$(((((a_0x+a_1)x+a_2)x+a_3)x+a_4)x+a_5)x$
12	加法	27	26	27	$y=(((((a_0x+a_1)x+a_2)x+a_3)x+a_4)x+a_5)x+a_6$

如果能保证自动改变加法指令第二地址内的单元番号，便能大大减缩计算程序。指令码保存在适当的单元内，并且可以把它们看成是一些数。在这些数上加以适当数时，便可以实现指令地址的自动改变。当利用这种方法时，多项式的计算程序将有表4的形式。

表4 多项式的计算程序

指令番号	指令码			
	运算码	第一地址	第二地址	第三地址
1	加法	20	—	27
2	乘法	27	31	27
3	加法	27	21	27
4	加法	3	28	3
5	比较	3	29	2
6	停止	—	—	—

第一指令是用来把第 20 单元的数转送至第 27 单元，以使得到标准化的乘法指令．完成第 2，3 指令后，即得值 a_0x+a_1．为要做下一步计算必须事先把加法指令(第 3 指令)的第二地址加 1，这正是由第 4 指令来执行的．按照这个指令选取位于第 3 单元的数，即选取正好需要的加法指令(第 3 指令)，并且在这个数上加以存储在第 28 单元内的量．为使得第 3 指令内的第二地址加一，则必须在第 28 单元内存储一个量：

运算码	第一地址	第二地址	第三地址
—	—	1	—

这样一来，在完成第 4 指令后就得第 3 指令的下列值：

运算码	第一地址	第二地址	第三地址
加法	27	22	27

新得的值存储在第 3 单元内，代替加法指令的原有值．

得到加法指令的新值后，便可以从乘法运算，即第 2 指令开始重复计算．为此目的我们有第 5 比较指令，这个指令将第 3 单元内的新指令与存储在第 29 单元内的量相比较．第 29 单元内保存如下的量：

运算码	第一地址	第二地址	第三地址
加法	27	27	27

这个比较最初给出，第一个量(第 3 单元中的)小于第二个量(在第 29 单元中的)，于是计算过程转至在比较指令第三地址中所示的第 2 指令．因此将自动地重复乘法指令(第 2 指令)与加法指令(第 3 指令)，并且每当执行一次加法指令第二地址内的单元番号便加一(由指令 4 来完成)．

这个循环的重复将一直进行到加法指令(第 3 指令)的第二地址达到 27 为止，因而须进行六次循环．此时第 3 指令有下列形式：

运算码	第一地址	第二地址	第三地址
加法	27	27	27

即指令码将与第 29 单元内所有的数相同．此时比较指令(第五指

令）将察觉到，保存在第3与第29单元内的量相等，于是计算过程转移到按番号顺序的下一条指令上，即第6指令．到此多项式的计算便告结束．

以程序来实现一些指令地址中的单元番号的自动改变方法在解各种各样问题时被广泛地采用着．它与循环重复方法相同，都能使我们有可能以相当少数的指令来实现极其大量的计算．

4．除了我们所考虑的三地址指令系统外，在许多计算机里还采用单地址指令系统．在采用单地址系统时，每一条指令内除了运算码外，只含有一个地址．为要对两个数执行运算以及把所得结果送至存储器，则需三条指令：这一条指令从存储器中选取一个数，并送至运算器，第二条指令选取另一个数，并对此二数执行所规定的运算，第三条指令将所得结果送至存储器．在计算时所得的结果往往是要用在下一次算术运算中的，在此情况下便无须将所得结果送至存储器，因而在作下一次运算时无须选取第一个数．因此在采用一地址指令系统时，程序的指令数量几乎比三地址的指令只多一倍．由于一地址指令所需要的位数比三地址指令的位数少，所以程序在存储器中所占用的位置在两种指令系统中大致相等（通常在采用一地址指令系统时存储器的每一个单元内保存两条指令）．在比较机器的工作速度时必须考虑到指令系统的不同特点．当以相同的速度执行运算时，一地址的计算机所执行计算的速度约比三地址的计算机慢一倍．

除了上述两种系统之外，有些计算机采用二地址及四地址的指令系统．

5．通常问题的解分为几个阶段，其中很多阶段对一系列的问题而言是共同的．例如，按给定变元计算初等函数，在解常微分方程时按给定导数确定函数增量，按已算出的被积函数来计算定积分等．

显然，对这种典型的阶段最好一劳永逸地制定好标准子程序．如果在解题进程中需要进行标准计算，则应在适当的瞬间把计算转移至一种标准子程序上．在标准子程序结束后必须回到基本程

序的中断处。

由于有了标准子程序便大大减轻了程序设计的负担。由于这种标准子程序的累积(它们记录在穿孔卡片或磁带上)许多问题的程序设计便可化为编制不长的连结个别标准子程序的基本程序.

6. 用电子计算机解决需要数亿次算术运算的问题. 只要在一个运算中出现错误,便会导致不正确的结果 显然,在这样大量的计算中进行人工检查实际上是不可能的. 因此检查的职能必须由机器本身来实现. 有一些用仪器检查运算的正确性的方法, 并且在发生错误时, 这种方法可以自动地停止机器. 但是这些仪器检查方法会导致仪器的大大增多, 而且通常不能包括机器的一切环节. 由计算程序本身直接规定的检查方法是比较有发展前途的

其中的一种检查方法就是在人工计算时所采用的重复计算——"双手计算". 如果在单独重复计算时得到相同的结果, 则足可以证明没有偶然性错误. 自然, 当有系统性的错误时利用这种方法是不能察出的. 为要免除系统性错误, 在解题之前用机器进行答案为已知的检查计算. 这些计算应该涉及到机器的一切环节, 在检查计算中所得结果正确时, 则可证明没有系统性的错误.

除了"双手计算"外, 根据问题的类型还可以采用更复杂的检查方法,例如在计算炮弹的弹道时,除了对相对速度的两个分量解微分方程组外,还可以额外地求解整个速度的微分方程,并在每一积分段上按公式

$$v^2=v_x^2+v_y^2$$

进行检查. 当解常微分方程时, 除了对积分步长 h 进行计算外还可以对步长 $\frac{h}{2}$ 进行第二次计算. 计算时这不仅能保证没有偶然性的错误, 并且也能对积分段选择的正确性予以估价. 在按递推公式计算表时, 有时可以用其它方法来计算出一些关节值. 关节值的正确出现足以保证算出的中间值的正确性. 在一系列的情况

中可按所得的结果之间的差来进行检查.

在编制程序时，必须规定对于已进行的计算这种或那种形式的逻辑检查.

数和指令的编码 数和指令在机器内用代码的形式表示. 在大多数的场合下采用二进位计数制，而不用一般的十进位制.

在十进位制中以 10 为计数制的基数. 每一位的数字可以取十个不同的数值：从 0 到 9. 次一位的单位值为原位的十倍，因此在十进位制中整数部分可以写为下列形式

$$N_{10}=k_0 10^0+k_1 10^1+k_2 10^2+\cdots+k_n 10^n,$$

这里 k_0, k_1, $\cdots$, k_n 可以取从 0 到 9 中的任一值.

在二进位制中以 2 为计数制的基数，每一位的数字只可以取两个数码：0 或 1，次一位单位值为原位的二倍. 因此在二进位制中整数部分可以写成

$$N_2=k_0 2^0+k_1 2^1+\cdots+k_p 2^p,$$

这里 k_0, k_1, $\cdots$, k_p 可取值 0 或 1.

在二进位制和十进位制中自然级数的前几个数可以写为下列形式

二进位制 0 1 10 11 100 101 110 111 1000 1001 1010 1011 等等

十进位制 0 1 2 3 4 5 6 7 8 9 10 11 等等

非整数部分利用基数的负幂按上述方式写出，例如 $3^1/_8$ 在二进位制中写成

11.001.

数从一种进位制转变为另一种进位制是按一定的算术运算进行的，并通常是在电子计算机上按特定的程序直接实现的.

对二进位数做的算术运算是与十进位制中的运算相同. 在二进位制中，当某一位的两个 1 相加时，在原位上得 0，而向上一位进 1. 例如，

$$1010+111=10001.$$

在二进位中乘法和除法比在十进位制中简单，因为乘法表被简单

的乘法规则代替了，也就是只用0或1去乘某数。例如，

$$
\begin{array}{r l}
1010 & (10) \\
\times\ 101 & (5) \\
\hline
1010 & \\
0000 & \\
1010 & \\
\hline
110010 & (50)
\end{array}
$$

在大多数计算机内采取二进位计数制的原因在于利用这种计数制可以大大地简化运算器(主要是在于执行乘法和除法运算)，除此之外，利用接通的或断开的继电器，在某线路上有或无信号的形式来表示每一位的数字是很方便的(在二进位制中每一位数字只可能取两种值：0或1)。

二进位数的每一位可以用电路上有无信号或继电器的状态的形式来表示，在此情况下每一位必须有自己的电路或继电器(图5)，而它们的个数应等于位数(并行制)。二进位数也可以用时脉冲码的形式来表示。在此情况下数的每一位隔一定的时间间隔沿一条线路进入(串行制)。通过每一位的时间标志是由整个机器所共用的同步脉冲来判定的。

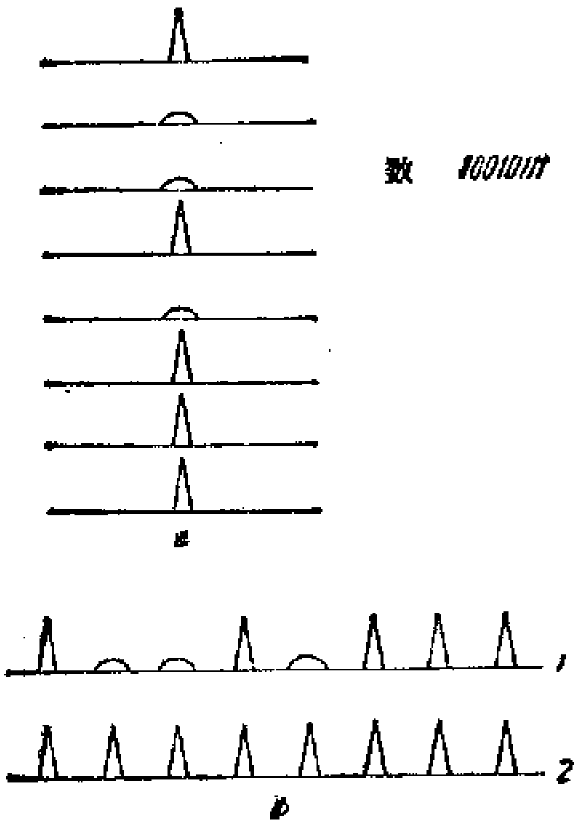


图5 代码的系统
(a)——并行的；(b)——串行的。
(1——代码；2——同步脉冲)

按照数的这些编码原理电子计算机分为两种类型：并行式计算机与串行式计算机。在并行式计算机中数的所有位同时传送，并且每一位都要有自己的电路。在串行式计算机中数的传送是沿一条电路进行的，但传送时间与位数

成正比，因此并行式计算机较串行式计算的速度快，但并行式计算机却需要较多的设备。

每一架电子计算机都具有一定的位数。在执行计算时所涉及到的一切数都不得超过一定的位数，自然这时就应该考虑到分割整数与小数的小数点的位置。

在一些计算机中小数点的位置是硬性选定的——这是定点制的计算机。通常小数点固定在第一个最高位之前，即一切数在计算中均应小于 1，这点是用选取适当的比例方法来保证的。在复杂计算中事先难以确定所得结果的位数，因此选择比例时应估计留出一定的保险位，但这样便降低了准确性，或者也可以在计算程序中规定自动变比例法，但这会使程序设计复杂化。

在某些计算机中对每一个数都指出小数点的位置——这是记阶法的计算机，或者通常称为“浮点制”计算机。指出小数点的位置就等于把数表为数值部分和阶的部分，例如：

$$N_{10}=10^k N'_{10} \quad 在十进位制中,$$
$$N_2=2^p N'_2 \quad 在二进位制中.$$

于是，数 97.35 可以表示为 $10^2\cdot0.9735$，表示一个数时在机器上应指出数的阶（p 或 k）和它的数值部分。这时数值部分的所有位全部被利用，而与数的大小无关，也就是任意一个数都用全部有效数字来表示，并带有同样的相对误差，这样便能提高计算的准确性，特别在执行乘法运算的时候更为突出，并且在大多数的场合下可以避免特殊选择比例。

在记阶法的计算机中计算准确性的提高和程序设计的根本简化的代价是运算器复杂化，主要的是执行加法和减法的部分。因为数可能有不同的阶，所以在做加法和减法之前必须先对位，并且在对位时舍去较小数的低位数，例如：

$$10^2\cdot0.7587+10^0\cdot0.3743=10^2\cdot0.7587+10^2\cdot0.0037$$
$$=10^2\cdot0.7624.$$

为适于二进位编码系统在定点制的计算机中数（它小于 1）只是用一连串的二进位符号来编码；例如：

$$.00110110000000=\frac{27}{128}.$$

在浮点制的计算机中，代码的一定部分用来记录阶，它也以二进位的记录形式编成代码．例如，数可写成下列形式：

$$6\frac{3}{4}=2^3\cdot\frac{27}{32}=0011.1101100000.$$

此外，为要记录代数符号(例如"+"表示为0或"−"表示为1)通常用一个阶的符号位和一个数的符号位．

指令与数都编成代码，代码的一定部分用来记录运算的番号(以二进位制的形式记录)和每一个地址的单元番号．

§3．快速计算机部件的技术原理在电子计算机上执行运算的次序

在计算机上按已取得的顺序指令执行每一个算术运算系分为如下阶段(假定我们考虑的是三地址指令系统)．

1．从存储器选取第一个数，并把它送至运算器(放置此数的存储单元番号在指令码的第一地址中指出)．

2．从存储器选取第二个数并把它送至运算器(单元番号在指令码的第二地址中指出)．

3．运算器根据运算码执行数的给定运算．

4．所得结果由运算器送至存储器的相应单元(单元番号在指令码的第三地址中指出)．

5．从存储器选取下一个相继指令，于是机器着手执行下一运算．

在机器内指令码送至"指令存储部件"(图6БЗК)．电子换接器(ЭК)把运算码的二进位番号变换为控制电压，并从相应于给定算术运算的一条电路中送出．这控制电压通过控制器($УУ_п$)准备好对执行必要运算所需的机器电路．

为要选取第一个数，指令的第一地址(A_1)的码沿地址代码总

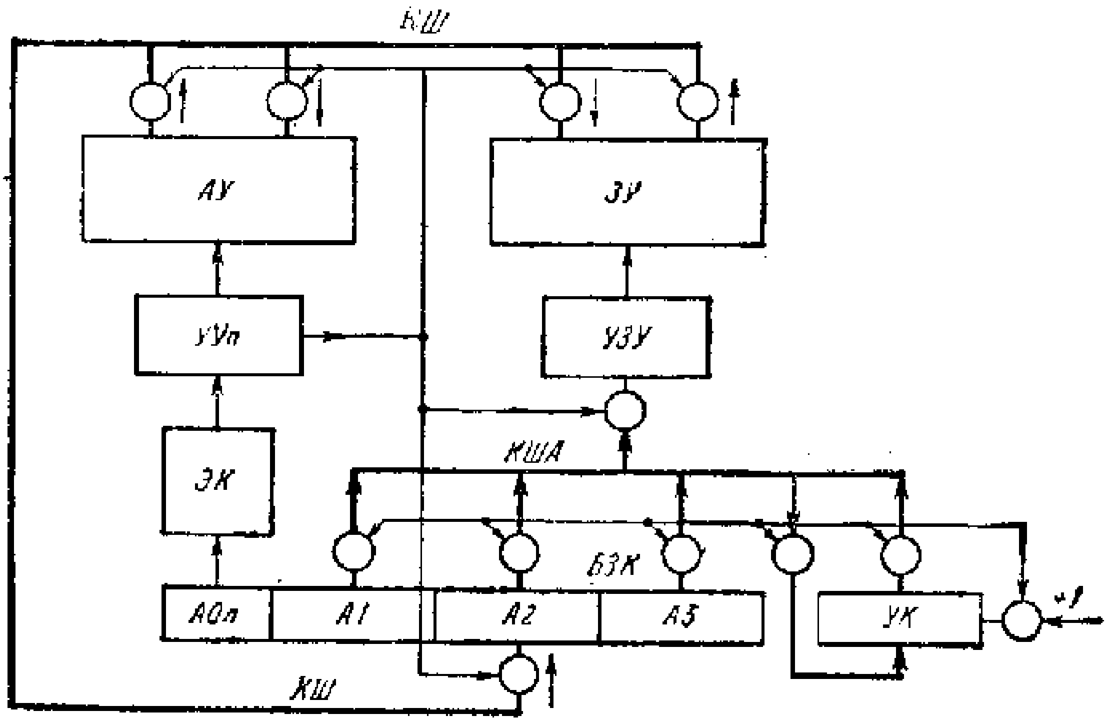


图 6　电子计算机结构线路图

线(КША)从指令存储部件(БЗК)传至存储器的控制部件(УЗУ).传送这个代码的信号是由机器的控制器(УУ$_п$)发出的. 第一个数是从相应于被传送的番号代码的存储器单元中选出的，然后它沿代码总线(КШ)进入运算器(АУ). 运算器的输入电路的开放是由机器的控制器(УУ$_п$)发出的相应信号来实现的.

第二个数也是以类似的方法选取的. 机器控制器(УУ$_п$)发出的信号把指令的第二地址($А_2$)的码从指令存储部件(БЗК)传至存储器的控制部件(УЗУ). 从存储器(ЗУ)中选出的第二个数沿代码总线(КШ)送至运算器(АУ).

运算器(АУ)根据预先已确定的运算码对这两个数进行给定的运算。

为要把所得结果传至存储器，指令的第三地址码($А_3$)沿地址代码总线(КША)从指令存储部件(БЗК)传至存储器的控制部件(УЗУ). 传送这个代码的信号是从机器控制器(УУ$_п$)送出的. 根据所接收的番号选取存储器的单元并将其输入电路开放. 选取数或接收数的状态是用机器控制器(УУ$_п$)所发出的信号来给出的.

由机器控制器($УУ_п$)所发出的信号将所得的结果从运算器(АУ)送到代码总线(КШ)上，数沿此条线进入存储器的已选定的单元.

为要选取指令在机器中规定有指令控制部件(УК). 在此部件中给出要选取的指令的番号. 通常指令是相继运行的，因此为要给出下一指令的番号必须在位于指令控制部件(УК)中的数上加1. 这点是由机器控制器来实现的(电路+1). 指令保存在存储器内. 为要选取下一指令新得的番号从指令控制部件(УК)沿地址代码总线(КШA)传至存储器的控制部件(УЗУ). 用来做传送的信号是由机器控制器($УУ_п$)发出的. 从存储器(ЗУ)中选出的新指令沿代码总线(КШ)传至指令存储部件(БЗК)，而这个部件的输入电路是由机器控制器的信号开启的，至此便完成了机器工作的一个循环. 在次一个循环里机器便执行新接受的指令. 指令按其番号的正常顺序可能在执行比较运算，例如在比较指令时遭到破坏. 这种指令不是执行某一算术运算，而是确定计算过程的进程. 如果第一数小于第二数，则必须转至执行在第三地址内所指出的番号的指令. 如果第一数大于或等于第二数，则选取下一条顺序的指令.

当比较指令代码进入指令存储部件(БЗК)时，电子换接器(ЭК)把运算码的二进位番号变换为控制电压并送至相应于该运算的一条输出电路上. 此控制电压准备好对执行比较运算所需要的机器电路.

从存储单元(在比较指令的第一和第二地址内指出的番号)选取两数是与在执行算术运算时完全相同的. 两数的比较可以在运算器(АУ)内用从第一数减去第二数的方法来实现. 根据结果的符号控制器($УУ_п$)或者把下一指令的番号码从第三地址(A_3)沿地址代码总线(КШA)送到指令控制部件(УК)上，或者对位于指令控制部件(电路+1)中的数加1，这点与算术运算相同. 在指令控制部件(УК)中确定出下一条指令的番号后，从存储器中选取指令，这点也与在执行算术运算时的选取相同.

运算器和控制器 电子计算机运用电子自动装置的现代工

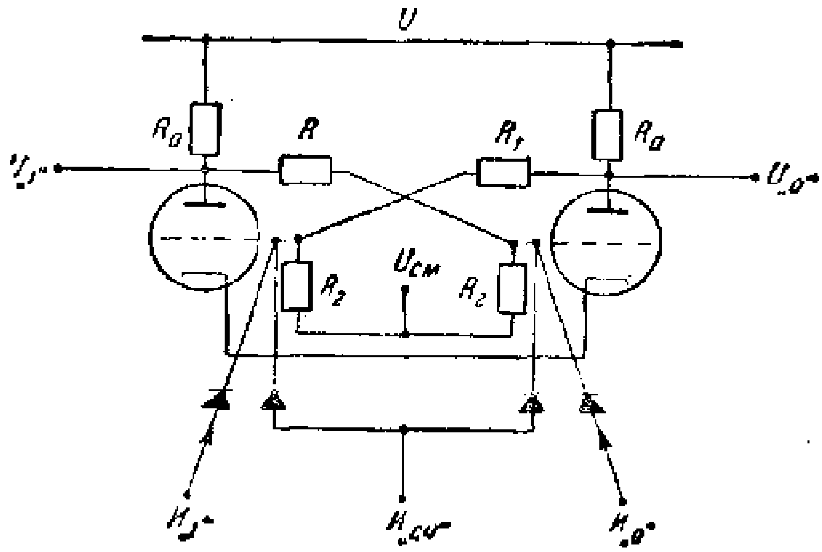


图7 触发单元线路图

具、基本上，计算机的各装置是按粗略的“是——非”原则，即是在实质上是有信号或无信号的原则来工作．正因为如此电子线路中的参数的相当大的变化也不会影响机器工作的正确性．

触发单元是电子计算机内所采用的最普通的元件之一．最普通的触发线路(图7)是两个带有阳极电阻 R_a 的放大器，它们通过分压器 R_1 和 R_2 相互作用着．利用固定偏压($U_{\text{см}}$)的方法来选择线路的这样一种状态，即其中一个电子管关闭，而另一个电子管开放．因为线路的两半部是对称的，所以任何一个电子管都可能被关闭，即线路有两种稳定的平衡状态．实际上，如果左边电子管关闭，而右边电子管开放，则在左边电子管($U_{\text{“1”}}$)的阳极上有高电压，而在右边电子管的阳极上有低电压(由于电子管电流引起在阳极电阻 R_a 上的低压缘故)．这些电压通过分压器 R_1 和 R_2 被送到另一个电子管的栅极上，于是在左边电子管的栅极上就有低电压，而在右边电子管的栅极上就有高电压．在正确选择线路的参数时，这些栅极电压可使电子管保持在给定的状态中．

同样，如果左边电子管开放，而右边电子管关闭，则在左边电子管的阳极上以及在右边电子管的栅极上就有低电压，而在右边电子管的阳极上以及在左边电子管的栅极上就有高电压．

触发单元从一种状态往另一种状态的转换可以通过二极管加

到电子管栅极上的负脉冲来实现．如果把负脉冲加到左边电子管的栅极上，则左边电子管就将截止，而它的阳极电压将增大，这样就会引起右边电子管的栅极电压加大，并以此开启右边电子管．因此，触发器处于第一个平衡状态（在左边的电子管的阳极上有高电压）．假如把负脉冲加到右边电子管的栅极上，则触发器处于另一个平衡状态（在右边电子管的阳极上有高电压）．如果把负脉冲同时加到两个电子管的栅极上，则每一个负脉冲都使触发器从一种平衡状态转换到另一种状态．

当我们把向电子管栅极传送脉冲的电路看做是线路输入端，而把阳极电压看做是输出端时，我们则得到触发单元的工作图表，如图 8．

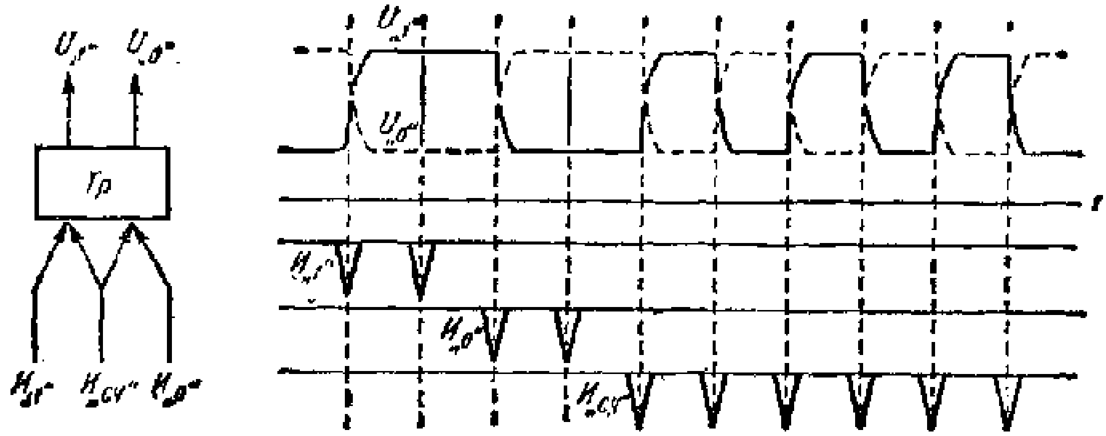


图 8　触发单元的工作图表

触发单元的特性适用于电子计算机的各种装置．触发器的一种平衡状态可以标明代码"0"，例如在右输出端（$U_{«0»}$）有高电压时；而另一种平衡状态可以标明代码"1"，例如在左输出端（$U_{«1»}$）有高电压时．于是输入端可以标明如下：$И_{«0»}$、$И_{«1»}$ 和 $И_{«сч»}$（计数输入端）．

在电子计算机内触发单元是用来暂时存储代码的（接收寄存器）（图 9）．用向所有单元的零的输入端上加负脉冲（$И_{«гаш»}$）的办法来事先将全部触发单元置于代码"0"的状态．数码或指令码以负脉冲的形式传送到触发单元的"1"的输入端．在具有代码脉冲的那些位上触发单元转换为代码"1"的状态，并将此状态一直保

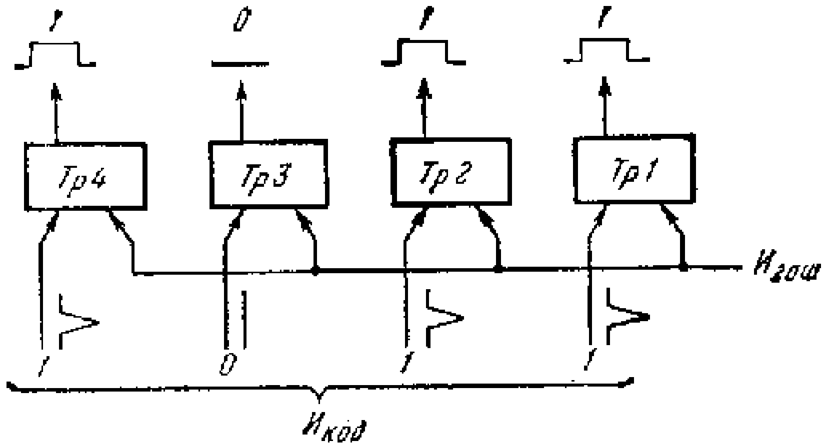


图 9　由触发单元组成的接收寄存器的线路

持到加入熄灭脉冲($И_{"гаш"}$)时为止. 在运算器内接收寄存器用来存储要执行的指令码,给出存储器所需单元的番号等等.

另一处采用触发单元的就是加法线路. 在这里每当把负脉冲加到计数输入端(同时加到两个输入端上)时, 利用触发单元的特性改变自己的平衡状态. 如果触发器处于代码"0"的状态时,则加入的脉冲使触发器转换为代码"1"的状态. 假如触发器处于代码"1"的状态时, 则加入的脉冲使它转换为代码"0"的状态. 当没有脉冲时,触发器就停留在原来状态. 触发器的初始状态可以看做是第一个数的给定位的代码, 而加入的脉冲可看做是第二个数的给定位的代码. 因此不难确信,触发器的性能完全对应于每一位的二进位加法规则(0+0=0; 0+1=1; 1+0=1; 1+1=10, 即"0"是在给定位上, 而"1"是向上一位进位). 为了实现二进位数的加法线路,必须保证从一位往另一位上进位. 在两个1相加时,即当触发器从代码"1"的状态转换到代码"0"的状态时, 产生向高位的进位. 在这种转换时触发器左输出端上的电压由高变低. 如果此电压被电容——电阻电路微分时, 则在电路的输出端上产生负脉冲. 此进位脉冲通过延迟线可以进入上一位的计数输入端.

图 10 是由触发单元组成的最简单的加法线路. 所有的触发单元都预先用脉冲 $И_{"0"}$(它加到零输入端上)置于代码"0"的状态. 在接收以负脉冲形式进入计数输入端的第一个数的代码时, 触发

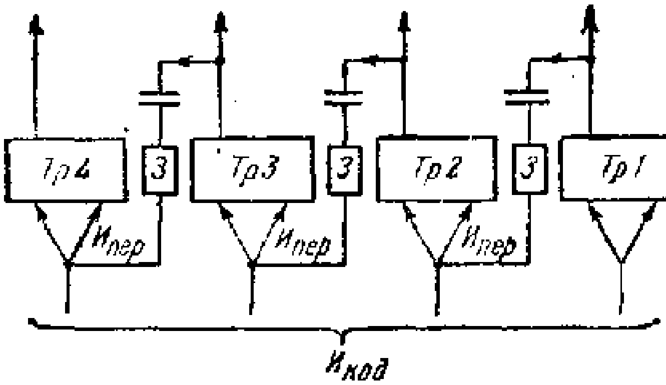


图 10　由触发单元组成的加法线路

单元将有相应于第一个数的代码的状态．在接收第二个数的代码时，就进行二进位数的逐位相加，并且在两个 1 相加的那些位上产生进位脉冲，它们经过延迟时间 t_0 进入高一位触发单元的计数输入端．这些进位脉冲可能使高一位的触发器从代码“1”的状态转换成代码“0”的状态．这时便又产生向更高一位进位的脉冲．在最坏的情况下，当代码相加时，所有的位都置于代码“1”的状态，而最低位从代码“1”的状态转换为代码“0”的状态，这时在每一位上都将产生进位脉冲，并且每个脉冲都要经过延迟时间 t_0．这样一来，进位脉冲所通过的总时间将等于每一位的延迟时间 (t_0) 乘上所有的位数．由触发单元构成的比较复杂的电子线路可以不采用类似的按位进位，因此就可以缩短加法的时间．

为了做数的乘法，由触发单元构成的运算器(图 11) 有两个用来存储被乘数和乘数 (P_1, P_2) 的接收寄存器和一个加法器 ($C_м$)．乘法按下列方式来实现．乘数的代码向右移一位，如果乘数的最后一位是代码“1”时，则在乘数寄存器的右输出端上产生脉冲，它给到控制电路(电路 $+u$)上，此电路可将代码从被乘数寄存器送到加法器上．然后在加法器中所取得的部分乘积向右移一位，并且重复上述运算．这样一来在加法器内将累积部分乘积的和．数码有多少位，这些运算就要重复多少次．在两个有“n”个位的数相乘时，乘积将有“$2n$”个位．乘积的高“n”位放置在加法器内．乘积的低“n”位可以在向右移位时放在乘数寄存器中所空出的位上．因此乘法结束时乘积的低“n”位留在乘数寄存器内．数码有多少位，

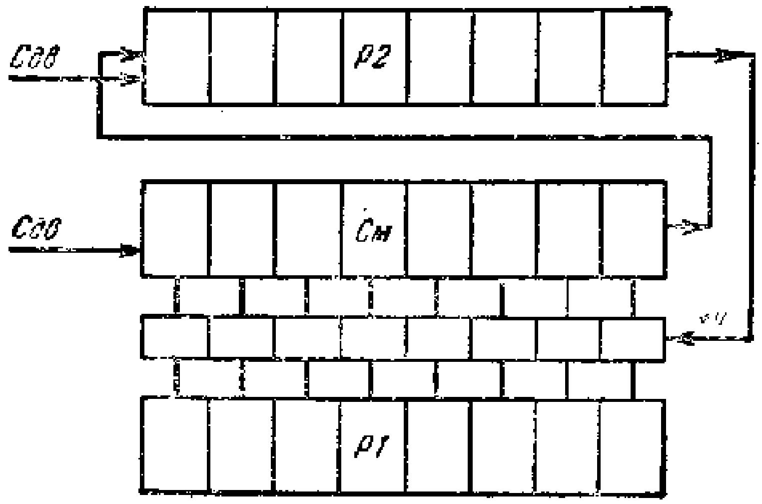


图 11　由触发单元构成的乘法线路

乘法时间大约就比加法时间多多少倍.

在触发单元内代码的移位按图 12 所表示的线路进行. 当把移位脉冲($И_{сдв}$)给到全部触发单元的零输入端时, 使它们置于代码"0"的状态. 在处于代码"1"的状态的那些触发单元内产生进位脉冲, 它们通过延迟时间 t_3, 使相邻单元置于代码"1"的状态. 因此每一个加入的移位脉冲都把代码移一位.

由触发单元构成的运算器也可实现数的除法. 这种运算器也是由两个接收寄存器和一个加法器组成的.

通常, 一个由触发单元构成的运算器对执行全部算术运算和逻辑运算来讲是作为通用的.

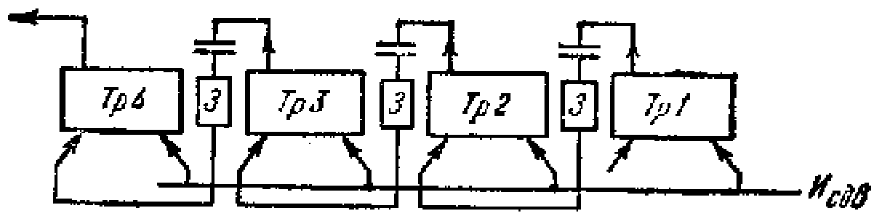


图 12　代码在触发单元内移位的线路

在电子计算机内触发单元也用来计算脉冲的数量, 这在一系列的控制线路中是必要的. 电子计数器线路(图 13) 不同于基本加法器线路(图 10)之处仅在于电子计数器的脉冲进位链内没有延迟

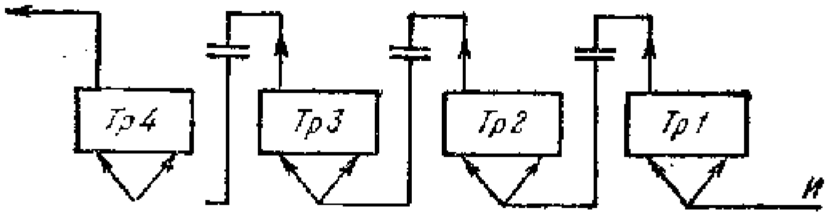


图 13　由触发单元构成的电子计数器线路

线，这样的计数器可计到 2^n 个脉冲（n——计数器的位数），在这之后计数器的状态重复．用某种使线路复杂化的办法可以取得计任意数量的脉冲的电子计数器（不等于 2^n）．

为要实现电子计算机的逻辑运算和控制线路采用了符合装置，反相器和二极管分离电路．

符合装置按逻辑定律"同——同"（即此又彼）工作，即只有在这种情况下，如果在全部输入端上都有信号时，在此装置的输出端才产生信号．反相器是按逻辑定律"是——非"工作的，即如果在输入端有信号，则在输出端就没有信号，并且相反，当在输入端没有信号时，则输出端有信号．二极管分离电路执行逻辑定律"或——或"，即只要在一个输入端上有信号的情况下，在输出端上就有信号．

在计算机内符合装置广泛地用来"选通"电气信号，即把信号送至必需的电路内．例如图 14 是数的一位的代码总线．存储器单元的输入端和输出端，运算器的两个接收寄存器的输入端及加法器的输出端都通过符合装置与代码总线相连接．当把控制信号给到存储器某个单元的输出符合装置上时，我们就把保存在此单元内的代码传送到代码总线上．如果于此同时把控

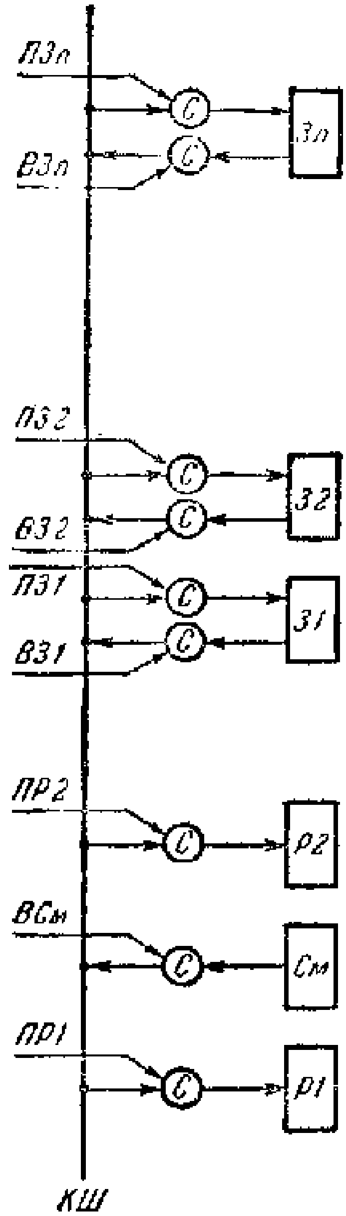


图 14　用符合装置控制信号的选通

制信号给到输入符合装置，例如第一个接收寄存器的输入符合装置时，则按着代码总线所传送的代码就进入第一个寄存器．同样，如果把控制信号给到加法器的输出符合装置时，则在加法器内所取得的代码传送到代码总线上．如果这时把控制信号给到存储器的某一单元的输入符合装置时，则按代码总线所传送的代码送到此单元去．当然，在把代码送到存储单元或运算器的接收寄存器时，必须事先消除里边原有的代码．

在电子计算机内用来做“选通”电气信号的符合装置的运用方法是很多的，只用上述的一个例子是不能详尽说明的．在存储器、运算器及机器的控制器内都广泛地采用符合装置．

表 5

代　码	第二触发器		第一触发器	
	左输出端	右输出端	左输出端	右输出端
“00”	H	B	H	B
“01”	H	B	B	H
“10”	B	H	H	B
“11”	B	H	B	H

H——在输出端上为低电压；　B——在输出端上为高电压．

除了执行“选通”信号任务外，符合装置能完成更复杂的职能．例如，在选择存储器的单元时常常需要把以二进位数的形式给出的单元番号转换为控制电压并且直接送到此单元内．这种任务是由以符合装置构成的电子换接器来实现的．图 15 是有四个输出电路的电子换接器线路．单元番号用两个触发单元以二进位代码的形式给出．这些触发单元的状态的四个可能的组合列入表格 5 内．

如果高电压对符合装置来讲是控制信号，则为要在“0”的输出电路内取得信号，必须把符合装置的输入端接到第一和第二触发器的右输出端上．在这种情况下，只有当触发单元处于代码“00”状态时，在此符合装置的输出端上才产生信号．同样，为了在第“1”输出端电路(代码“10”)上取得信号，相应的符合装置的输入端

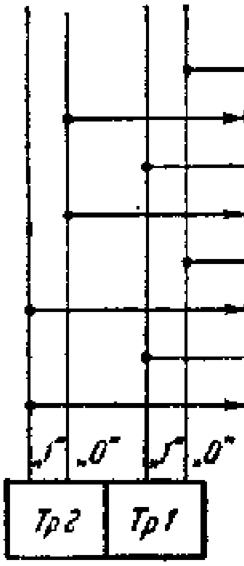


图 15 有四个输出电路的串子换接器线路

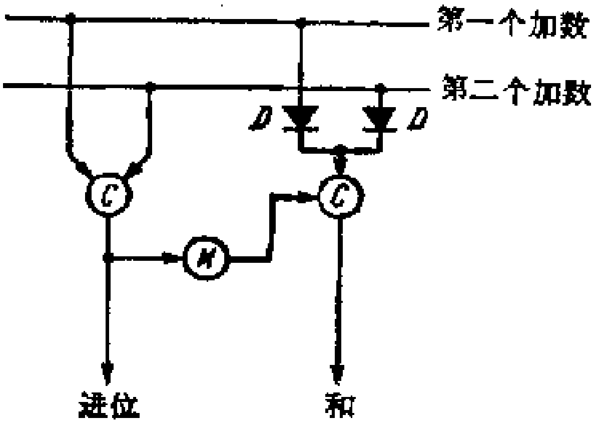


图 16 一位的半加法器线路

应当接到第一触发器的左输出端和第二触发器的右输出端上．对第“2”（代码“10”）和第“3”（代码“11”）电路也按此原则连接符合装置．

在很多情况下为要建立运算器常常同时采用符合装置，反相器和分离二极管．对于两个二进位的逐位加法同样也有四个可能的组合：

表 6

№	两个加数的数值		和的值	向高一位进位
	第一个数	第二个数		
1	0	0	0	0
2	0	1	1	0
3	1	0	1	0
4	1	1	0	1

这些关系的执行可以由图16所表示的线路来实现．这样的线路称为“半加法器”．向高位进位的信号由符合装置（组合 4）来进行．当输出端上没有进位信号时，为要取得和的信号（组合 2 和 3）只要在其中的一个输出端上有信号就够用了，这可由符合装置，反相器和二极管连接电路来执行．在两数相加时，除了给定位的

数字外，也必须考虑到由前一位转来的进位. 此进位可以这样来考虑，即是把从前一位传来的进位重复加到所取得的结果上. 这样一来，把两个半加法器串连起来便可完全保证两个二进位数的一位加法. 一位的加法器线路也可直接实现，只要在研究可能的组合时能考虑到由前一个低位传来的进位.

在顺序传送代码的计算机内采用由符合装置构成的加法线路最为合适. 在此情况下数码按一个代码总线传送. 数的所有位应当通过严格给定的一定时间间隔相继地进行. 在此场合下可采用一位的加法器(图 17). 两数的代码进到一位加法器的两个主要输入端(数是先从低位传送的). 进位的输出通过延迟线 (3) 进到加法器的第三输入端. 延迟时间等于脉冲之间的间隔时间. 于此，如果数的某些位相加时产生进位脉冲，则此脉冲恰好在下一个高位的脉冲到来的瞬间给到加法器的输入端上. 两数相加的时间等于一个数的代码的通过时间.

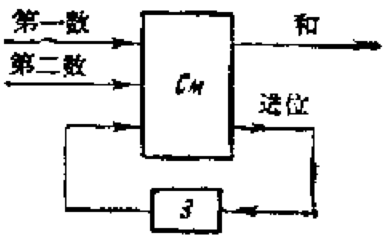


图 17 由符合装置构成的串行加法器线路

在采用顺序代码时在一位的加法器上也可进行两数的乘法. 这时，数码有多少位，数就必须在加法器内通过多少次，即乘法时间要比加法时间多"n"倍.

存储器 计算机的可能性在颇大程度上取决于存储器的容量，即在计算机内能够存储数的数量. 现代的通用电子计算机的容量一般为 500—4000 个数.

为了保存代码可利用触发单元. 但是这时的设备就非常庞大，因而几乎不采用这种存储器.

在串行式的计算机内电声汞管存储器得到了普遍应用(图 18). 电信号以脉冲形式进入管子输入端的石英晶体上. 石英晶体具有使电脉冲转换为机械振荡和相反的转换的性质. 这样一来，输入的电信号转换为机械振荡(超声波的)并使它沿着管子以一定速度传播. 当信号到达管子的末端和要进入石英的接收晶体时，它就

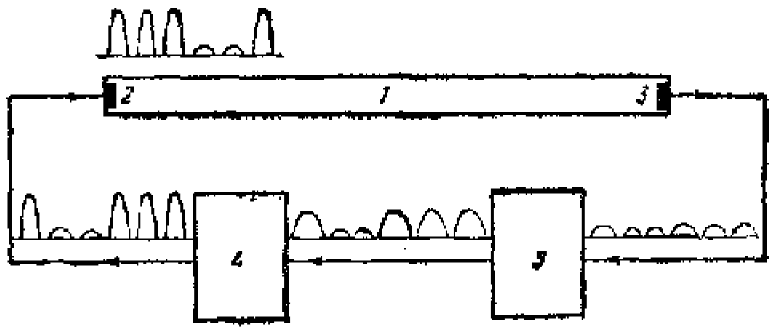


图 18　用电声管动态存储代码的原理线路

1——汞管；2——石英的传送晶体；3——石英的接收晶体；
4——恢复脉冲的形状；5——脉冲放大.

又转换为电脉冲．当信号经过放大及恢复它的原来形状时，从新把信号传送到管子的输入端．因此，以脉冲形式引入汞管内的数码将无限期地沿着管子循环．为了输入数，代码从机器内给到管子的输入端上，并且同时在此时间间隔切断管子末端的脉冲循环电路．为了在相应的瞬间，即当所需代码到达管子末端时选取数，就要开放能传送代码至机器其他装置的输出电路．数的输入和输出是由相应的电子线路自动实现的．通常为了减缩设备，在每一支汞管内存储几个数．因此，在选取数时必须等待所需的代码到达管子的末端为止．在管子内存储的数愈多，等待所需数的平均时间也愈长．

采用电声汞管做存储器的串行式计算机的工作速度每秒钟大约为一、二千次运算．

经常采用电信号的磁记录原理做存储器，这种原理类似于录音方法．记录或者用磁带，或者用镀有铁磁材料的不断旋转的磁鼓进行(图 19)．磁头是沿磁鼓的纵线放置的．如果在一定瞬间电流脉冲通过磁头的线圈，则信号就以剩磁形式被记录在磁鼓表面的相应位置上．当磁鼓旋转时，剩磁场在磁头下通过时，将使磁头感应出电信号，它们经放大器放大之后，送至机器的其他装置内．磁鼓既可用在代码的串行传送系统中，又可用在代码的并行传送系统中．但是，电声汞管的缺点——在选取数时的等待时间——不但是磁鼓所固有的，而且程度还很大．因此磁鼓存储器

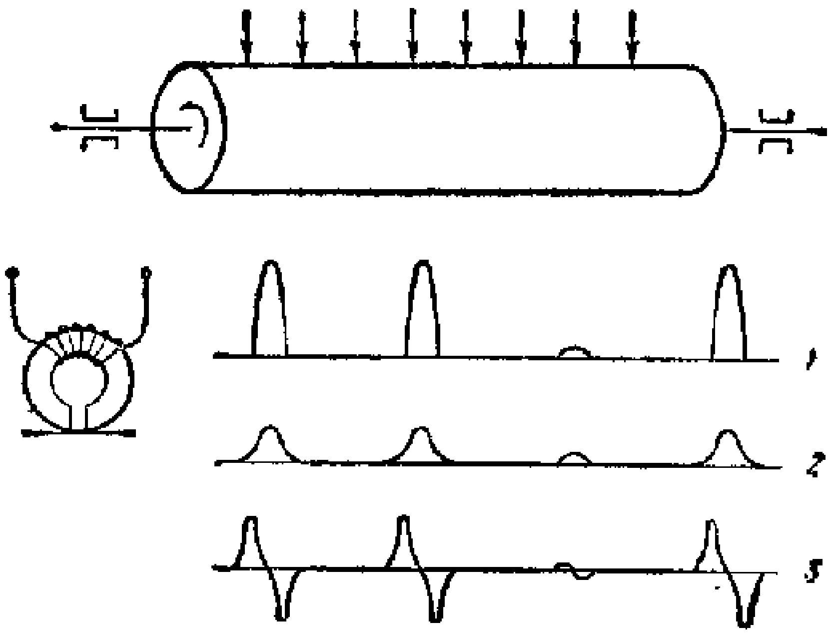


图 19 磁鼓的原理图

1——通过线圈的电流；2——剩磁；3——读数时在线圈内的电动势.

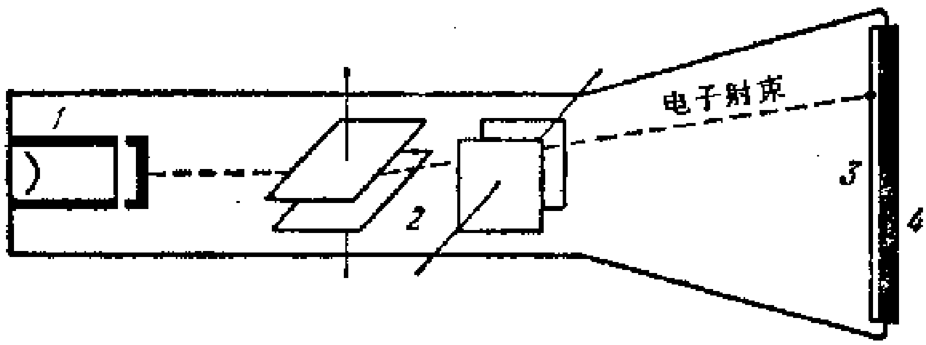


图 20 电子射线管的原理图

1——电子枪；2——偏转板；3——屏幕；4——信号板.

一般在速度较慢的计算机内(大约每秒钟为几百次运算)采用. 但是另一方面,磁鼓能够在少量的设备下大大增加存储器的容量,这在有时候有本质的意义. 因此在通用计算机内除了快速(内)存储器之外,时常采用磁鼓和磁带做补助(外)存储器.

在并行式的快速电子计算机内采用电子射线管(图 20)作存储器. 如果电子射束射到介质屏幕上的某一点上,则在此点上受到充电. 电荷将保存很久,因而在屏幕上可以记录数码. 读数时,电子射束重新射到所需的点上. 如果给定部分没有被充电,则此

部分就进行充电，并且有电流脉冲通过信号板和输出放大器．如果此部分已充电，则信号就不进入．因此可以确定，在给定点上是否记录着信号．在选取代码之后必须使此部分恢复原有状态，这点是用专门线路自动实现的．同样也必须周期地恢复所进行的代码记录，以便防止由寄生电子和介质漏电所引起的电荷的较大变化．

通常，在屏幕上容纳1024(32×32)或2048(32×64)个点．使电子射束射到所需的点上是用把相应电压加到两对偏转板上的办法来实现的．

在并行式计算机内二进位数的每一位都有自己的电子射线管，而数的选取是从所有的管子同时进行的．选取时间(包括整个线路的工作)可能达几微秒．

最近以来用具有矩型磁滞回线的磁性元件作成的存储器(图21)得到了采用．如果通过线圈加正信号，则磁心就受到正磁化，加负信号时，就受到负磁化．

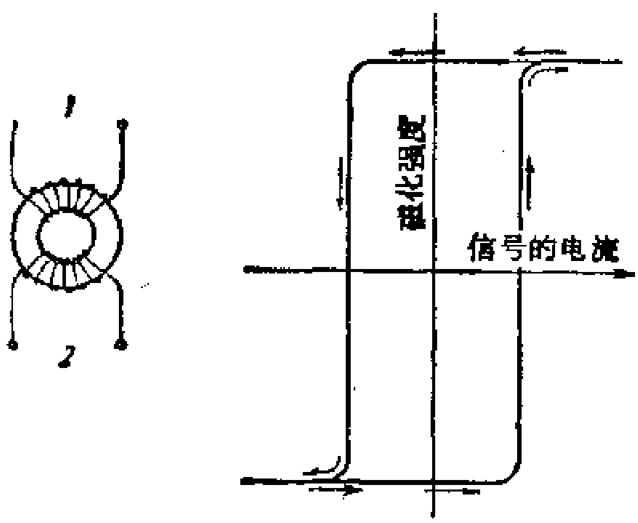


图21 具有矩型磁滞回线的存储元件的原理图
1——输入线圈；2——输出线圈．

当去掉信号时，磁心仍然保留在磁化状态，或者是正磁化，或者是负磁化．因此，磁心的状态可标明所记录的信号．在读数时，通过线圈加入一定极性的信号，例如正信号．如果这时磁心原为负磁化，则对它进行反磁化(改变磁通)并且在输出线圈上感应出电动势，然后它由放大器放大．如果磁心的磁化原是正的，则不改变它的状态，并且在输出线圈内不产生信号．于是，可以辨别出，在此元件上记录着甚么样的信号．自然在选取代码之后必须恢复磁心的原有状态，此工作是由专门线路实现的．

§4. 电子计算机的发展和使用的远景

电子计算机的应用对现代科学和技术的很多领域的发展，特别是对物理数学科学的发展应有巨大的影响. 因此试图指出计算机的进一步应用的基本前途及其对数学应起的作用是很适宜的.

数学机器运用领域的进一步扩展

1. 目前通过进一步改善结构，运用新的物理原理和新型部件在制造快速计算机的领域中有着不断的和重大的技术进步，因而可以期待这些计算机的技术特性(快速，存储量，工作中的连续性及可靠性)的改进以及它们的结构和操作的大大的简化和便利，这样便可保证这种计算机的广泛使用的可能性.

计算机的类型繁多也可促进计算机的广泛使用. 除了生产率很高的大型计算机外还有小型的，即是操作简便的，任何科学及设计研究所或工厂可以置备的小型计算机，除了通用计算机外还有用来解一定范围内的问题的比较简单的专用计算机；除了纯数字计算机外还制造一种这样的计算机，即它能按仪表上的指数不断接受数据和以数字形式加工数据以及以曲线或参数形式给出结果，以便控制与计算机相连接的装置.

2. 提高这些计算机的使用效率的另一个方法就是进一步的改进程序设计的方法. 用§2内所描述的一般方法只能对比较简单的数学问题编制程序；而编制大的实际问题的程序还是一件非常复杂而冗长的工作. 使用由对于计算基本函数和执行某些常用的数学运算(矩阵反演，数值积分)而一次编好的标准子程序所组成的"图书馆"在某种程度上可以减轻程序设计的工作. 虽然如此，基本程序与子程序的编排，结果的赋地址法和改地址，程序的检查和调整仍是一件需要一定技巧的相当复杂而冗长的工作. 这种情况会大大阻碍对电子计算机提交新的工作.

在此方向中有两条进一步发展的道路. 第一是利用计算机本身自动组织程序，即利用计算机按专门的"设计程序的程序"把以

代码形式输入到机器内的问题的基本公式和逻辑方案转变为程序.

第二种方法是：计算机在按某种特制的通用的程序操作时，根据已输入机器内的，含有一系列巨大算题(例如，解方程组)的总的计算方案不经编制详细的工作程序而来直接分析和执行运算，同时保证结果的正确放置和输出.

3. 计算机在数学方面应用中的进一步发展是与不仅在执行数值运算而且也在执行分析运算时，使用计算机是有关系的.

原则上，在很多情况下这种可能性是显而易见的，而且也是完全可以实现的. 比如说，如果多项式用它们的系数列来记录，则多项式的乘法和除法这一类的运算只是对于系数列的算术运算，在计算机上是很容易编制程序的. 利用确定的编码来记录函数时，完全可能编成按初等函数给出(用同一编码记录的)导数，即可以进行分析形式的微分. 这样今后就有可能按固定方法(例如用幂级数解微分方程组)实现解题并完全实现分析运算和数值运算. 因此，计算机可以用来执行相当细致和高水平的脑力工作(但只能是定型的工作)，正如普通机器用来不仅代替土工而且也代替刺绣工的劳动一样.

快速计算机对数值方法和近似方法的影响 当然，所采用的各种劳动手段和工具对工作方法本身有影响. 例如利用对数计算的三角公式在使用直接执行乘法和除法的计算机时是不合适的. 台式自动计算机的应用使得近似方法中的某些计算方案(例如，微分方程中的非差分方案)成为合适的.

因此，计算工具中的那些根本改变和由于电子计算机的运用而开辟的那些可能性不仅使我们对数值分析方法应给予重新的估价，而且在某种程度上更一般地对数学的问题及其应用也给予重新的估价.

我们列举几个这方面的问题，其变化对这些问题来讲是最清楚的.

数学表和引入函数进行计算的其他方法 首先电子计算机根

本上改变了数学表计算的可能性.从前每年只能制出一种函数表,而将来却可制出数百种,因而使我们有可能制出一个变数甚至于几个变数的所有基本特殊函数的完整而准确的表来.同时表的结构也应当有本质的改变.在应用快速计算机时采用含有稀疏基值而针对高阶内插的紧凑表是很方便的.

在很多场合下,利用其他引入函数的方法——最佳近似法的区间多项式,展开成连续小数,基于函数积分表示式的数值解近似公式等来代替表是更方便的,这些方法应当导至于给定函数的计算程序.

特殊函数和特殊的分析解 特殊函数的运用和最终以分析形式构成解对问题的定性研究,以及对阐明它们特性(在数值解时考虑这些特点也是很重要的)仍保留自己的作用.对某些大容量的问题它也可能就提供最经济的数值解法.但同时在利用结构上复杂的工具和特种函数来构成在许多特殊情况下为可能的准确解或近似解的办法(从前是为了减轻它的计算而这样做的)是没有充分根据的.在应用计算机时以利用一般的数值方法,而不用上述分析表达的可能性来求解可能是更为简便迅速.

因此,过去为取得技术学科和力学中的某些特殊问题的复杂的分析形式的解而花费的巨大努力在很多场合下将是不必要的了.

数值方法的选择 如果认为由于电子计算机有高度的生产率而无须发展近似方法并且可以应用最原始的近似方法,这是不正确的.实际上,只有在解极简单的一维问题才不依赖于方法的选择——运算的次数只有几十次或者几万次,——这样的问题在电子计算机上只用几秒或几分钟就可解完.

因为在系统地解新的,比较复杂的问题时,一个问题常常都需要执行几千万和几亿次的运算,所以正确的选择能够缩减运算次数的方法也是极其重要的,因此研究有效的近似方法首先对多维的问题(多变数函数的补插,多重积分的计算,非线性代数和超越方程组的解,空间积分方程的解,偏微分方程及方程组的解等)是

有着现实意义的.

同时对数值方法的评价的看法也大大地改变了：方法是应当根据在机器内实现它们的可能性和通用性，即按其应用的宽度和大量性来评价. 因此那些基于问题的局部特点和计算者的技巧的方法在一定程度上失掉了价值.

能应用于各种各样问题的通用方法，例如差分法，变分法，梯度法，迭代法，线性化法等等）应当具有最大意义.

当然，在选择数值方法及实现它们的方案时应当注意到这种方法是要用机器来实现的，并且有时也应当考虑到机器的结构特性. 特别是最大限度地利用内存储器的可能性，减少从外部输入的数据量，执行中间检查的可能性，问题的程序设计的方便性更为重要.

但是不应当认为，在机器内只能实现以同一类型运算为基础的最简单的方法. 程序设计的广泛可能性及其方法的进一步改善能使我们在计算机上实现非常复杂的带有各种分支和按所得结果改变计算过程的计算方案，这些方案用人工计算也是难以实现的. 主要的只是在程序中都应完全规定出所有这些可能性.

同样也不应该认为，在机器内不能实现需要代数运算的方法. 上面已经提到，执行某些分析运算也是完全可能的.

误差估值的意义 在近似方法的误差估计问题中，渐近特性的估计应有很大的意义，因为在差分法中大的值 n（例如用代数方程组代替积分方程时的方程个数），小的步长等在快速电子计算机上都是完全可以实现的. 据此，在对近似方法做比较评价时，表明方法收敛速度的特点的渐近值会取得决定性作用.

当在机器上实现计算方法时，误差的后天估计——根据算出的解的估计——应有更大的应用. 这种估计可以被包括在计算程序之内；计算进程可能依赖于它们的结果. 比如，如果误差大于容许限度时，把步长减小一半自动重复计算. 在这方面后天估计会比先天估计要更加方便和更加现实. 先天估值常常偏高，并且结构也远为复杂.

问题的理论分析的可能性 应当指出另一种利用由问题数值解所得的数据的可能性。这就是在运用函数分析的方法时，按所取得的近似解应用泛函分析的方法可以判断解的存在性和统一性，同时也可以确定解的分布区域。由于用纯理论的方法来进行这种研究有时是非常复杂和冗长的，并且在某些特殊问题中这种研究实际上是不可能实现的，因此，利用这种为此目的在机器上进行的数值计算的可能性无疑的是有趣的。

数值方法中的新问题 计算可能性的急剧增大及其运用实践经验的积累在数值方法的研究中引起了新问题。在过去解含有大量未知数的线性方程组的问题只是个别的，而现在这种方程组在解数学问题时已成为经常的了，因而关于不仅由系数引起的，而且也由计算过程引起的舍入误差对于未知数的精确度的影响问题有了现实意义。对这个问题已进行了一系列有意义的研究。

在机器上对于大区间上微分方程组应用小步长数值积分的可能性使得关于方程的数值积分过程的稳定性问题更加显著了。此问题的试验性的分析和随后所进行的理论研究导致对微分方程的数值积分方法的本质上的重新评价。

稳定性的问题对偏微分方程差分解法也具有极重要的意义。

新方法 计算机使用的可能性使得出现了完全新式的近似和数值方法，或使得从前被认为是完全不能实现的方法成为完全可以现实的。一个典型例子就是随机试验法(或通常称为"蒙特-卡罗法)。这方法在于对求我们所要定的未知量找出一个概率问题，此问题的解(概率，数学期望)要与我们的未知数量相符合。这个概率问题的解通过随机试验实验地求得，它表为一系列试验的平均值。例如，为要确定由不等式 $F(x, y) \geqslant 0$ 决定的位于正方形(0.1;0.1)之内的图形的面积，则在正方形中选择随机数偶(x, y)，而确定它们满足于上述不等式的比较。当然，如果用人工进行这些试验的话，这种方法的效率是非常低的，但是，如果利用计算机，则它就完全可以实现。试验本身则可借助于随机数的表来实现。对某些问题，例如以不大的准确度求多维积分，这种方法可能比其他

方法更加有效.

这类方法可以用来解矩阵的反演问题，如果采用适当的马尔科夫网路，也可以用来解偏微分方程，如果已指出相关的随机过程.

快速计算机对于数学分析、力学和物理学的作用 在数学分析中，有关数学物理中的积分方程和边界问题的多维问题的研究具有本质上较大的兴趣和实际的意义. 这些研究和已找到的解法将不再是无益处的，它们很可能由于利用计算技术的新工具得到完全实现，尤其是，这种问题的系统解现已具有极其现实的意义了.

自然,新研究出的解法应该考虑到实现它们的可能性.

另一方面,在进行数学问题的初步研究时,由于利用计算机以足够的准确度进行多次的试验计算的可能性大大地扩展了"数学实验"的应用范围并提高其效率. 因此总的来说,以及在各别问题中,例如在微分方程的定性研究中,应用数学实验方法的研究成为很重要的了.

有趣味地也应指出，在数学分析问题中可以使用计算机不仅是为了联系到实用,而且也为了它的内在理论的需要. 例如,用计算机进行计算对于确定某些函数不等式和估值中的常数会是很必要的;机器计算不仅在分析中,而且在数论中都取得了这种应用.

最后,计算机可以用来检查数理逻辑公式的真伪性,而且由于一系列的数学命题和证明可以用数理逻辑的记号写出，因而在原则上有可能用快速计算机检查某些数学推理的逻辑的正确性.

如果谈到力学和物理学时，则首先应该着重指出数学在这两门科学中的应用的巨大的扩展. 直到目前为止数学在数学物理的具体问题中的运用是受到必要的计算的复杂性和工作量所限制的. 在实际问题中这种范围通常是这样的，只是算一个问题就需要几个月,有时甚至需要几年的计算工作. 因此,尽管力学和理论物理学的许多问题的一般数学提法已经明确，并且在原则上已研究出了一系列的解法，但实际上准确的或数值的数学解只能在为

数不多的某些理想化了的和简化了的场合下才能实现(平面或轴对称的问题、最简单的边界回线、无限长的机翼等).

因此，数学解与其说是用来寻求所需计算的量还不如说是用来对问题进行定性或定向的研究，这种研究在实践上必须以昂贵的实验来补充.

但是，新式计算工具的应用为按照真实情况的复杂程度广泛地解决力学和物理学问题开辟了可能性(如空间问题、带有复杂的边界回线问题、非线性方程、偏微分方程组等).

当然，为要实现这一可能性要求进一步研究上述问题的数值分析方法和用计算机解它们的方法. 但是利用快速计算机成功地解气象学和气体动力学中的偏微分方程组，颗粒介质方程及其他问题的现有经验已经可以证明这种可能性是可以实现的.

对实现力学和物理学问题的理论数学分析以及使这些问题的提法接近于实际物理问题的条件的广泛可能性，和由于采用快速计算机所赋予此分析法的快速性和灵活性使得我们有可能在很多的情况下以数学实验代替物理实验. 这样一来，将使物理和力学问题的研究方法有进一步的改善以及使理论方法和计算方法在这种分析中占据更重要的地位.

电子计算机对技术和生产的作用 数学分析的问题的快速性和有效的数值解也可以使技术问题的理论研究法在各种技术学科(建筑力学、电工学和无线电技术、水力学和水力工程学等)中广泛地利用，并且使这种分析更加准确和现实. 这样我们也将有可能把数学分析应用在直到如今尚未用过这种分析的那些技术问题中去.

除了数学分析的问题(在技术科学中常遇到的问题)的数值解外，数学机器在技术中的运用还有全然不同的其它的可能性. 例如在技术设计中，在选择设计决议的方案或者选择某些设计项目的配置时，都可以采用数学机器. 在组织生产的问题中，选择工作的分配和次序时，可能有各种不同的解答方案. 选择最好的、生产率最高的和最经济的方案是一件非常困难的事情. 于是机器就在这儿

也找到了用处，因为如果能对上述方案的系统的反复考虑过程设计出程序，并且能算出我们所需要的特性时，则利用机器计算就可以比较几万和几十万个方案，这在用一般方法时是力所不及的.

尤其是，很多种继电器一接触线路能以数理逻辑方法进行分析和检查，这种数理逻辑方法是可以用快速计算机来实现的. 用这种方法可以按这些或那些特征来比较上述线路的某些方案，并从中选出最好的方案.

在生产的自动管理方面利用计算机是非常有前途的，只要是把这些机器与跟踪仪和传送器联接在一起时. 例如，根据送入机器内的有关加工品的几何数据，机器可按一定程序制造产品，并给出确定机床工作状态的参数和必要的改变. 由于运算速度快，同一架电子计算机可用来同时控制几台机床的工作. 这类计算机在自动控制运动的物体中，例如控制星际导弹，所具有的意义是十分清楚的. 因为控制程序除考虑预先引入的数据外还能考虑到由仪表指数所表示的情况的变化.

这样一来，关于计算机的构造和工作分析及其应用它们的可能性的问题对数学来说是一个广泛的活动领域. 最近的将来数学机器的使用对苏联技术和文化的发展必然会起很大的作用.

文　　献

С. А. Лебедев, Электронные вычислительные машины, Изд. АН СССР, 1956.

Ф. В. Майоров, Электронные цифровые вычислительные машины, "Природа", № 11, 1954.

С. Л. Соболев, А. И. Китов и А. А. Ляпунов, Основные черты кибернетики, "Вопросы философии", № 4, 1955.

М. Уилкс и др, Составление программ для электронных счетных машин, ИЛ 1953.

用一架计算机的例子阐述程序设计的一般原理.

Быстродействующие вычислительные машины, Сборник переводов под ред. Д. Ю. Панова, ИЛ, 1952.

А. И. 基多夫，电子数学计算机，科学出版社，1958 年版.

于桂芝　译

冯　康　校

第三卷

数学名著译丛

数 学

它的内容，方法和意义

〔俄〕A．D．亚历山大洛夫 等 著

科学出版社
www.sciencep.com

内 容 简 介

本书是苏联数学界普及数学知识的一部名著．全书共二十章，分三卷出版．书中不仅综述了现代数学的各分支学科，还综述了数学的哲学、历史发展及其在物理和工程技术方面的应用．本书每章都由苏联第一流学者撰著，有很高的学术水平．同时，本书写得深入浅出，只要具备高中的数学知识，就可以理解．在努力实现四个现代化的今天，这部书必将帮助我国广大读者扩大眼界，找到解决问题的数学工具．

本卷共六章，内容包括实变数函数论、线性代数、抽象空间、拓扑学、泛函分析、群及其他代数系统．

本书可供大学数学系师生、中学教师、具有高中以上文化程度的学生、有关方面的科技人员及数学爱好者阅读．

А. Д. Александров

МАТЕМАТИКА, ЕЁ СОДЕРЖАНИЕ, МЕТОДЫ И ЗНАЧЕНИЕ/ТОМ 3/

Издательство Академии Наук СССР, 1956

数 学

——它的内容、方法和意义——

第三卷

〔苏〕А. Д. 亚历山大洛夫 等 著

王 元 万哲先 裘光明 孙以丰

田方增 刘绍学 吴品三 王隽骧 译

沈信耀 李培信 江嘉禾

秦元勋 校

科学出版社 出版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1962年5月第一版 开本：850×1168 1/32

1984年4月第二次印刷 印张：10 5/8

印数：9,901—36,000 字数：282,000

统一书号：13031·2495

本社书号：3428·13—1

定价：1.60 元

目　　录

第　三　卷

第十五章　实变数函数论(C. Б. 斯捷奇金著)……………………1

§1. 绪论 ……………………………………………………1

§2. 集合论 …………………………………………………2

§3. 实数 ……………………………………………………10

§4. 点集 ……………………………………………………16

§5. 集合的测度 ……………………………………………24

§6. 勒贝格积分 ……………………………………………30

第十六章　线性代数(Д. К. 法德杰也夫著)……………………35

§1. 线性代数的对象和它的工具 …………………………35

§2. 线性空间 ………………………………………………46

§3. 线性方程组 ……………………………………………58

§4. 线性变换 ………………………………………………72

§5. 二次型 …………………………………………………81

§6. 矩阵函数和它的一些应用 ……………………………89

第十七章　抽象空间(A. Д. 亚历山大洛夫著)…………………93

§1. 欧几里得公设的历史 …………………………………93

§2. 罗巴切夫斯基的解答 …………………………………97

§3. 罗巴切夫斯基几何 ……………………………………103

§4. 罗巴切夫斯基几何的现实意义 ………………………112

§5. 几何公理，它们利用一定的模型来检验 ……………120

§6. 从欧几里得几何分出的独立的几何理论 ……………127

§7. 多维空间 ………………………………………………134

§8. 几何对象的推广 ………………………………………148

§9. 黎曼几何 ………………………………………………160

§10. 抽象几何和现实空间 …………………………………173

第十八章　拓扑学(П. С. 亚历山大洛夫著)……………………186

§1. 拓扑学的对象 …………………………………………186

§2. 曲面 ……189
§3. 流形 ……194
§4. 组合方法 ……197
§5. 向量场 ……205
§6. 拓扑学的发展 ……210
§7. 度量空间与拓扑空间 ……213
第十九章　泛函分析(И. M. 盖尔芳特著) ……218
§1. n维空间 ……219
§2. 希尔伯特空间(无穷维空间) ……222
§3. 依直交函数系的分解 ……228
§4. 积分方程 ……234
§5. 线性运算子及泛函分析进一步的发展 ……241
第二十章　群及其他代数系统(A. И. 马尔采夫著) ……251
§1. 引言 ……251
§2. 对称和变换 ……252
§3. 变换群 ……260
§4. 费得洛夫群 ……272
§5. 伽罗华群 ……280
§6. 一般群论的基本概念 ……283
§7. 连续群 ……292
§8. 基本群 ……294
§9. 群的表示与指标(特征标) ……301
§10. 一般群论 ……306
§11. 超复数 ……306
§12. 结合代数 ……316
§13. 李代数 ……324
§14. 环 ……327
§15. 格 ……333
§16. 一般代数系统 ……335

第十五章　实变数函数论

§1. 绪　　论

在十八世纪末、十九世纪初的时候，微积分学已基本上被发掘了．在这个时候（正确地说是十八世纪），学者们开始建立它的各个分支，揭示了很多新而又新的事实，发展了微积分学在力学、天文学与技术科学方面的种种问题的日新月异的应用．现在已经出现综观所有获得的结果的可能性，使它们系统化并深入探究分析的基本概念的意义．但在这个时候，已经可以看出来，分析基础本身并不是毫无问题的．

还在十八世纪，那时的大数学家关于什么是函数就没有一致的见解，因此长期争论着问题的这样与那样的解答，这样与那样的具体数学结果，弄不清究竟谁是正确的．逐渐才知道了分析的另一些基本概念也需要进一步精确．如果对什么是连续性及连续函数的性质是些什么，都没有足够清晰的理解，那就常常会引起一系列错误的论断．例如，认为连续函数总是可微的．由于数学中已经处理着这样复杂的函数，以至于仅仅依凭直观与猜测已经是不可能的了．因此，分析的基本概念的整理就成为燃眉之急了．

第一个在这方面认真尝试的是拉格朗日，在他之后，柯西又走上了这条道路，柯西将直到今天还广泛运用的极限、连续性与积分的定义严格化了．大约就在那个时候，捷克数学家波尔察诺也认真地研究着连续函数的基本性质．

我们详细地讨论一下连续函数的一些性质．命 $f(x)$ 是某个间隔 $[a, b]$（即满足不等式 $a\leqslant x\leqslant b$ 的所有的点 x）上的连续函数．如果在间隔的端点上，函数取异号的数值，那么在间隔中间必定有一点，使函数的值为零．这件事在过去认为是显然的，现在这个命

题已经得到了严格的论证. 同样严格地证明了间隔上的连续函数必定在其上的某些点取自己的最大值与最小值.

连续函数的这些性质的研究, 迫使人们深入地去探究实数的本性, 因此就出现了实数的理论, 而数轴的基本性质也得到了确切的表述.

数学分析的进一步发展, 促使人们必须去研究愈来愈多的"坏"的函数, 特别是间断的函数. 例如, 间断函数作为连续函数的极限而出现. 早先, 人们并不知道, 极限函数是否连续. 间断函数也在描绘突然改变的过程时出现, 因此产生了一个新的问题——把分析的工具推广到间断函数上去.

黎曼研究了这样一个问题, 即积分的概念可能推广到怎样的间断函数上去. 由于所有这些奠定分析基础的活动, 使一个新的数学分支——实变数函数论出现了.

如果古典数学分析是基于处理一些"好"的函数(例如连续函数和可微函数), 那么实变数函数论主要是处理更为广泛的函数类. 如果在数学分析里对于连续函数给出了某些运算(例如积分)的定义, 那么实变数函数论就以研究这样一些问题为特征, 即这个定义可以应用于怎样的函数类, 以及这个定义应该怎样改变, 才能更为广泛. 特别是, 只有在实变数函数论里才能满意地回答这样的问题, 即什么是曲线的长度? 及对于怎样的曲线, 讨论它的长度才有意义?

实变数函数论本身是建立在集合论的基础上的.

正由于这个原因, 我们先叙述集合论的原理, 然后转入点集的讨论, 而以叙述实变数函数论的基本概念之一——勒贝格积分作为本章的结束.

§2. 集 合 论

人们经常需要考虑各种各样事物的集合, 正如第一章(第一卷)所阐明的, 这就引起了数的概念的产生, 然后是集合的概念的产

生；这是数学中基本而又简单的概念之一，不能再加以精确的定义了．以下讨论的目的，无非是要弄清楚集合是什么，而不必纠缠于它的定义．

不管按照什么特征或者依循什么规律结合起来的事物的总体都称之为集合．集合的概念是通过抽象化的途径而产生的．人们把任意东西的总和看成集合，是抽去了集合中各个东西之间的所有联系与关系，而仅仅保留了这些东西的个别特性．这样一来，由五个钱币做成的集合与五个苹果做成的集合就完全不相同了．但是另一方面，围成一个圆圈的五个钱币做成的集合与一个个叠起来放的五个钱币所做成的集合，则被看成是相同的．

我们试举几个集合的例子，例如，沙粒的集合，太阳系所有行星的集合，在给定的时间里，在某个房子里所有人的集合及这本书的全部书页的集合．数学里也常常碰到各种各样的集合，例如，给定的方程所有的根的集合，全体自然数的集合及直线上所有的点的集合等等．

研究集合不依赖于组成它的事物的特性的性质，即仅仅研究集合的一般性质的数学分支称之为集合论．这一分支是在十九世纪末及二十世纪初才开始蓬勃发展起来的．德国数学家康托尔(G. Cantor)是科学的集合论的奠基人．

康托尔关于集合论的工作产生于三角级数收敛性问题的研究．由于具体数学问题的研究而引起非常抽象与普遍的理论的建立，已经是一种非常普通的现象了．这些抽象理论的意义在于它不仅与产生它的那个具体问题有联系，而且由于它对于另外一系列问题也有着应用．集合论也就是如此．集合的想法与概念已经浸透到所有的数学分支，并且改变了它们的面貌，所以不熟悉集合论的原理就不可能对近代数学获得正确的理解．而集合论对于实变数函数论则有着特别巨大的作用．

假如对于任何东西，都可以知道它属于集合或者不属于集合，那么集合就算被给出来了；换言之，集合是由所有属于它的东西所完全确定的．如果集合 M 是由而且仅仅是由 a, b, c, …这些东

西构成，那么就写成

$$M=\{a,\ b,\ c,\ \cdots\}.$$

组成任意集合的东西，通常称为它的元素．当一个东西 m 是集合 M 的元素时，就记作

$$m\in M.$$

读做“m 属于 M”或者“m 是 M 的元素”．如果有一个东西 n 不属于集合 M，则记作 $n\overline{\in}M$．每一个东西只能是给定的集合的一个元素；换言之，一个集合中所有的元素彼此都是不同的．

集合 M 的元素，本身可以是集合．但是，为了避免矛盾起见，应该要求集合 M 本身不是组成它自己的一个元素，即 $M\overline{\in}M$．

不包含任何东西的集合称为**空集**．例如，方程

$$x^2+1=0$$

所有的实根组成的集合就是空集．以后我们用 θ 来表示空集．

对于两个集合 M 与 N，如果集合 M 的每个元素 x 同样也是集合 N 的元素，那么就说 M 含在 N 之中；M 是 N 的一部分；M 是 N 的子集或者 N 包含 M．而记作

$$M\subseteq N \quad 或 \quad N\supseteq M.$$

例如集合 $M=\{1,\ 2\}$ 是集合 $N=\{1,\ 2,\ 3\}$ 的一部分．

显然 $M\subseteq N$ 恒成立．为了方便起见，我们把空集看作是任何集合的子集．

如果两个集合是由完全相同的元素构成的，那么就说这两个集合相等．例如，方程 $x^2-3x+2=0$ 所有的根的集合与集合 $\{1,2\}$ 是相等的．

我们来定义集合上的运算规律．

联或和　给出集合 $M,\ N,\ P,\ \cdots$．集合

$$X=M+N+P+\cdots,$$

即由它的“被加数”$M, N, P, \cdots$ 的所有元素所组成的集合，称为这些集合的联或和．但是，如果元素 x 属于很多被加数，那么 x 只能归入和 X 一次．显然

$$M+M=M.$$

而且当 $M \subseteq N$ 时，有

$$M+N=N.$$

交 同时属于所有的集合 $M, N, P, \cdots$ 的那些元素的全体所组成的集合 Y，称为 $M, N, P, \cdots$ 的交或这些集合的公共部分。

显然 $M \cdot M=M$. 如果 $M \subseteq N$，则 $M \cdot N=M$.

如果集合 M 与 N 的交是空集，即 $M \cdot N=0$，则称这两个集合不相交。

为了标记集合的和与交的运算，我们仍旧沿用记号 $\sum$ 与 $\prod$. 因此

$$E=\sum E_i$$

为诸集合 E_i 之和，而

$$F=\prod E_i$$

为它们的交。

请读者自己证明，集合的和与交服从普通的结合律

$$M(N+P)=MN+MP.$$

同样也服从规律

$$M+NP=(M+N)(M+P).$$

差 所有属于 M 而不属于 N 的元素所组成的集合 Z，称为两个集合 M 与 N 的差：

$$Z=M-N.$$

如果 $N \subseteq M$，则差 $Z=M-N$ 也称为集合N 关于 M 的补集。

不难证明，关系式

$$M(N-P)=MN-MP$$

与

$$(M-N)+MN=M$$

恒成立。

因此，集合的运算规则与普通算术运算规则有很大的不同的.

有限集合与无限集合 由有限多个元素组成的集合，称为有限集合. 如果集合的元素的个数是无限的，那么这个集合就称为无限集合. 例如全体自然数所组成的集合就是无限集合.

我们研究任意两个集合 M 与 N, 并且提出这样的问题，即这两个集合的元素的数量是否一样.

如果集合 M 是有限的，那么它的元素的数量可以由某个自然数(即其元素的数目)来表达. 在这种情形之下，为了比较集合 M 与 N 的数量，就只要计算一下 M 与 N 的元素的个数，然后比较一下所得到的这两个数目就可以了. 同样，假若集合 M 与 N 中，一个是有限的，另一个是无限的，那么很自然地可以认为无限集合包含着比有限集合更多的元素.

然而，如果两个集合 M 与 N 都是无限集合，那么用简单地计算元素的个数的方法是什么也得不到的. 所以立刻引起这样的问题，即是否所有的无限集合的元素的数量都是一样的，或者是否存在元素数量互相不同的无限集合？ 假如后者是正确的，那么用什么方法可以比较无限集合的元素数量呢？我们就要讨论这些问题.

一一对应 重新命 M 与 N 为两个有限集合. 如果不利用计算集合中元素的个数的方法，又如何来判断这两个集合中哪一个包含的元素更多一些呢？ 为此我们将要确定它的对应关系，即将 M 中的一个元素与 N 中的一个元素联结成一对. 因此，假若 M 有任何一个元素在 N 中找不到它所对应的元素，那么显然 M 的元素较 N 为多. 我们只要观察一些例子，就很清楚了.

假若大厅里有若干个人及若干把椅子. 为了要知道哪样多一些，只要让人们都去找座位. 如果有人没有位子坐，那么人就比椅子多. 但是，假如所有的位子都有人坐上了，那么人和椅子正好一样多. 上面所讲的比较集合元素的数量的方法比依次计算元素个数的方法有着无比的优越性. 这是因为不需要特别的改变，这个方法就不仅能够用之于有限集合，而且可以用之于无限集合.

我们研究所有自然数的集合

$$M=\{1,\ 2,\ 3,\ 4,\ \cdots\}$$

与所有偶数的集合

$$N=\{2,\ 4,\ 6,\ 8,\ \cdots\}.$$

哪一个集合包含着更多的元素呢？一眼看去，似乎前者的元素更多些．然而，我们可以在这两个集合的元素之间建立一对对的关系，如下表所示：

表 1

M	1	2	3	4	…
N	2	4	6	8	…

没有任何 M 的元素，也没有任何 N 的元素找不到它所对应的元素．我们也可以建立如下的对应关系：

表 2

M	1	2	3	4	5	…
N	—	2	—	4	—	…

因此 M 有很多元素找不到它所对应的元素．但是另一方面，我们也可以将对应关系确定如下表：

表 3

M	—	1	—	2	—	3	—	…
N	2	4	6	8	10	12	14	…

现在 M 又有很多元素没有对应的元素了．

因此，如果集合 A 与 B 是无限的，那么用不同的方法来建立对应关系，就得到完全不同的结果．假如存在这样一个建立对应关系的方法，使 A 的每个元素与 B 的每个元素都有它所对应的元素，那么就说在集合 A 与 B 之间可以建立一一对应的关系．例如，恰如表 1 所示，上面所考虑过的集合 M 与 N 之间，就可以建立一一对应的关系．

如果在集合 A 与 B 之间可以建立一一对应的关系，那么就说他们的元素有同一数量，或者称它们同势．如果用任何方法来建立对应关系，集合 A 中总有若干个元素没有与之对应的元素，那

么就说，集合 A 的元素比 B 多，或者 A 的势比 B 大。

因此，我们获得了上面提出的问题的解答，即如何比较无限集合的元素的数量。然而这却一点也不能给另外一个问题以丝毫的回答，即是否存在不同势的无限集合？为了要回答这一问题，我们来研究某些简单类型的无限集合。

可数无限集合 如果可以在集合 A 的元素与全体自然数所组成的集合

$$Z=\{1, 2, 3, \cdots\}$$

的元素之间建立一一对应的关系，那么就称集合 A 是可数无限的。换言之，假若集合 A 的元素可以用全体自然数来标记号码，即可以将它写成序列的形式

$$a_1, a_2, \cdots, a_n, \cdots,$$

那么就说 A 是可数无限集合。

表 1 说明了全体偶数所组成的集合是可数无限的（前一列的数，现在被看成是它所对应的后一列各数的指标数）。

由上所述容易看出，最小的无限集合正是可数无限集合，即每一无限集合皆包含一个可数无限集合为其子集。

如果两个非空的有限集合互不相交，那么它们的和的元素的个数比它们中的任何一个的元素个数都多，但对于无限集合，这个规律就可能不成立。事实上，命 $Ч$ 是所有偶数所组成的集合，H 是所有奇数做成的集合，而 Z 是全体自然数组成的集合。表 4 说明，集合 $Ч$ 与 H 都是可数无限集合。然而 $Z=Ч+H$ 仍然是可数无限集合。

表 4

$Ч$	2	4	6	8	…
H	1	3	5	7	…
Z	1	2	3	4	…

对于无限集合，"整体多于部分"这一法则被破坏了，这就表明无限集合有本质上异于有限集合的特性。从有限过渡到无限，完

全符合了熟知的辩证法的规律——性质的质变.

现在我们来证明全体有理数所组成的集合是可数无限集合. 为此我们将有理数排列成下表:

(1)	(2)	(3)	(4)	(5)	(6)	
1	2	3	4	5	6	…
0	-1	-2	-3	-4	-5	…
$\frac{1}{2}$	$\frac{3}{2}$	$\frac{5}{2}$	$\frac{7}{2}$	$\frac{9}{2}$	$\frac{11}{2}$	…
$-\frac{1}{2}$	$-\frac{3}{2}$	$-\frac{5}{2}$	$-\frac{7}{2}$	$-\frac{9}{2}$	$-\frac{11}{2}$	…
$\frac{1}{3}$	$\frac{2}{3}$	$\frac{4}{3}$	$\frac{5}{3}$	$\frac{7}{3}$	$\frac{8}{3}$	…
$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{4}{3}$	$-\frac{5}{3}$	$-\frac{7}{3}$	$-\frac{8}{3}$	…
…	…	…	…	…	…	…

此处第一列按照大小递升的次序安排了所有的自然数, 第二列遵循递降的顺序安置着零及负整数, 第三列又依照递升的次序安放着分母为 2 的所有正的既约分数, 第四列则又遵循着递降的顺序排列着分母为 2 的所有负的既约分数, 如此等等. 显然, 每一有理数在这张表上都能够而且只能被找到一次, 现在我们将这张表上所有的数都按箭头指示的方向, 依次标以号码. 于是, 所有的有理数都依次被安排成一个序列如下:

有理数对应的指标数	1	2	3	4	5	6	7	8	9	…
有　理　数	1	2	0	3	-1	$\frac{1}{2}$	4	-2	$\frac{3}{2}$	…

这样一来, 在全体有理数与所有的自然数之间建立了一一对应, 于是, 全体有理数的集合是可数无限的.

连续统势的集合 假如可以在集合 M 的元素与间隔 $0\leqslant x\leqslant 1$

的全体点之间建立一一对应的关系，那么就称集合 M 具有连续统势。例如按照这个定义可知，间隔 $0\leqslant x\leqslant 1$ 的点的集合本身，就具有连续统势。

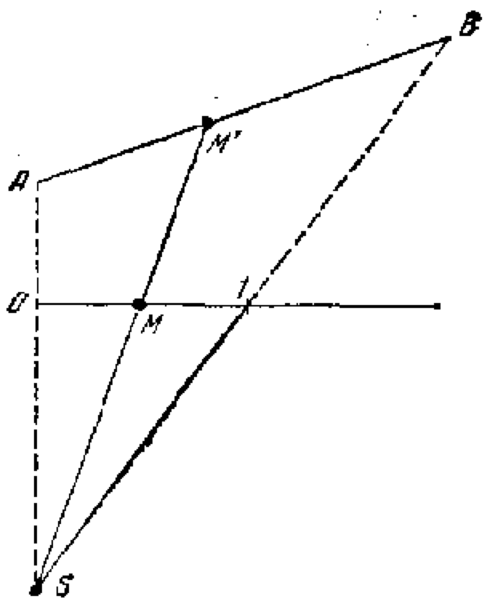


图 1

由图 1 可以看出，任何间隔 AB 的点的全体具有连续统势。由几何投影的方法，建立了它的点与间隔 $0\leqslant x\leqslant 1$ 的点之间的一一对应关系。

不难证明，任何区间 $a<x<b$ 与直线 $-\infty<x<+\infty$ 上所有点所组成的集合，都具有连续统势。

更为有趣的事情是正方形 $0\leqslant x\leqslant 1$, $0\leqslant y\leqslant 1$ 的点的集合有连续统势。因此，粗略地说，正方形中的点与间隔上的点"一样多"。

§3. 实　　数[1]

数的概念的发展已经在第一章(第一卷)中详细介绍过了，现在我们将粗略地介绍实数的理论。这个理论是由于要建立分析的基本概念，在十九世纪产生的。

有理数　我们假定读者已经熟悉有理数的基本性质。这里我们只回忆一下这些性质，而不去详细叙述它们。有理数集合是由形如 m/n 的数所做成的数集，此处 m, n 为整数，而 $n\neq 0$。有理数有两个由一些规律(公理)所确定的运算(加法与乘法)。以下用 a, b, c, … 来表示有理数。

I. 加法公理

1) $a+b=b+a$　(交换律)。

2) $a+(b+c)=(a+b)+c$　(结合律)。

1) 写这段时，作者与柯尔莫果洛夫(А. Н. Колмогоров)作了有益的讨论。

3) 方程 $$a+x=b$$

有唯一的解(存在逆运算).

由这些公理立刻可以看出，表达式 $a+b+c$ 有确定的意义. 存在有理数零(零元素)，使 $a+0=a$. 又存在加法的逆运算——减法，因而表达式 $b-a$ 是有意义的.

从代数的观点来看，对于加法的运算关系，全体有理数做成交换群.

II. 乘法公理

1) $ab=ba$ (交换律).

2) $a(bc)=(ab)c$ (结合律).

3) 方程 $$ay=b$$

当 $a\neq0$ 时有唯一的解(存在逆运算).

由这些公理立即得知，表达式 abc 是有意义的，存在有理数 1 使 $a\cdot1=a$. 又对于异于零的有理数，存在逆运算——除法. 因此对于乘法运算来说，除零外的所有有理数做成交换群.

III. 分配律

1) $(a+b)c=ac+bc$.

有理数满足所有的公理 I—III. 对于加法及乘法的运算关系来说，全体有理数组成一个所谓代数域.

IV. 良序公理

1) 对于任意两个有理数 a 与 b，在下面三个关系中，有一个而且仅有一个成立：$a<b$，$a>b$ 或 $a=b$.

2) 若 $a<b$，$b<c$，则 $a<c$.

3) 若 $a<b$，则 $a+c<b+c$ (加法的不变性).

4) 若 $a<b$ 与 $c>0$，则 $ac<bc$ (乘以 $c>0$ 的不变性).

由于所有这些公理都满足，所以我们称有理数集合为良序域. 除有理数外，还有其他事物的集合也满足这些公理，因而也是良序域.

我们还应该注意下面所讲的有理数的两个重要性质：

稠密性 对于任何有理数 a 与 b，而 $a<b$，总可以找到有理

数 c，使得 $a<c<b$.

可数性 全体有理数的集合是可数无限集合(见 § 2).

量的度量 仅仅考虑量的度量这样一个重要的问题，就可以看出，对于数学的发展来说，仅仅有有理数是不足的. 我们观察一个最简单的例子，即间隔长度的度量问题.

考虑一条直线. 在直线上确定了方向、原点（点 O）及单位度量. 这样一来，就确定了端点在 $\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $-\frac{2}{3}$ 等等的间隔 OA. 一般说来，每一个有理数 a 都在直线上对应一个点 A，即点 A 的坐标 $x=a$. 因此 a 就表示方向为 OA 的间隔的长度. 然而用这个方法所定义的长度，并不是每一个间隔都是可以用某个(有理)数来度量的. 例如，在古希腊时就已经熟知的，边长为单位的正方形的对角线的长度，就不能用任何有理数来度量. 换言之，直线上的点多于有理点. 这就很自然地引起了这样的问题——建立数与直线上的点的一一对应关系，也就是进一步扩充数的概念.

实数 我们已经说过，仅用有理数去度量量是不够的，因而数的概念应当进一步扩充，使数与直线上的点一一对应起来. 为了这个目的，我们希望弄清楚，仅仅借助于有理点是否可能确定直线上任意点的位置，从而类似于有理数的构造，使我们得到实数的概念.

命 α 为直线上的任意点，于是所有的有理点 a 可以分成两类：一类是位于 α 左方的全体有理点 a，另一类则是位于 α 右方的全体有理点 a. 如果这两类有一交点 α(假如 α 正好是有理数)，那么 α 可以属于任何一类. 有理点的这种分类就称为分割. 如果两个分割的左方与右方的有理点都是相同的(至多除掉一点)，那么就称这两个分割恒等. 现在不难看出，不同的点 α 与 β，所确定的分割也是不同的. 事实上，由于有理数是处处稠密的，所以总可以找出两个有理数 r_1 与 r_2，都位于 α 与 β 之间且它们都不等于 α 或 β，那么对于一个分割，它们位于其右方，对于另一个，就位于其左方.

因此直线上每一点都确定了有理数范围内的一个分割．不同的点对应不同的分割．值得重视的是分割也可以用与此多少有些不同的方法来定义，也就是让 α 本身不出现在定义内．把有理数的全体分成两个非空非交的集合 A 与 B，对任意 $a\in A$，$b\in B$ 都有 $a<b$，我们就称这是有理数的一个分割．这样定义的分割，唯一地确定了一个点(界点)．换言之，借助于有理数范围内的分割可以确定直线上的任何点．这种构造方法是德国数学家戴德金首创的，因此后人称之为戴德金分割．

分割并不是借助于有理点来确定直线上任意点的位置的唯一可能的方法．普通度量所常用的是下面要讲的康托尔的方法．重新命 α 为直线上任何一点，那么总可以找到两个有理数 a 与 b，其距离小于任意给定的常数，使 α 位于 a 与 b 之间．这样一来，a 与 b 就确定了 α 的近似位置．我们介绍的这种近似确定点 α 的位置的过程可以不断延续下去，使每作一次，精确度就增加一次．因此得到了端点为有理数的一系列的间隔 $[a_n, b_n]$，满足 $[a_{n+1}, b_{n+1}]\subseteq[a_n, b_n]$ 及 $b_n-a_n\to0(n\to\infty)$．满足这些条件的诸间隔，称为间隔套．这样的间隔套唯一地确定了点 α 的位置．

借助于有理数类似的构造，也可以定义实数．进一步，我们将定义实数上的运算关系，使之满足与有理数上的运算关系同样的公理．现在由于直线上的每一点都对应一个实数，且其逆也真，所以常常称实数的全体为数轴．

连续性准则　有理数全体所成的集合与全体实数所成的集合之间存在着本质的差别，那就是实数的全体具有一系列的性质，它们刻划出这个集合的连续特性．但有理数的全体所成的集合却不具有这些性质．这些性质就称为连续性准则．现在我们列举其中重要的几条．

戴德金准则　若将所有的实数分成两个无公共元素的非空集合 X 与 Y，对于任意的 $x\in X$，$y\in Y$，皆满足 $x<y$，那么存在唯一的数(界)ξ，对于任何$x\in X$，$y\in Y$ 皆有 $x\leqslant\xi\leqslant y$．适合 $a\leqslant x\leqslant b$ 的实数 x 的全体，称为直线上的间隔，我们记之为 $[a, b]$．若

$[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ 及 $b_n - a_n \to 0 (n \to \infty)$，则称间隔系 $[a_n, b_n]$ 为间隔套.

康托尔准则 对于任何间隔套 $[a_n, b_n]$，皆存在一个而且只存在一个点 ξ，它属于每一个间隔 $[a_n, b_n]$.

魏尔斯特拉斯准则 每一递增有上界的实数序列都收敛.

若对于任何 $\varepsilon>0$，都可以找到自然数 N，对于所有的 $n>N$ 及所有的自然数 p 皆有

$$|x_{n+p} - x_n| < \varepsilon,$$

则称实数序列 $\{x_n\}$ 为基本序列.

柯西准则 实数的每一基本序列都是收敛的.

由于我们并未严整地建立起实数来，因而，我们在此还不能证明这些准则对于实数是正确的. 我们的中心目的在于研究这些准则彼此之间的关系是怎样的. 假定对于实数，连续性准则中有一个是正确的，我们来研究如何由此推出其余的准则.

我们来讨论所有的连续性准则都是等价的这样一个一般的命题.

假如：1) 对于所有的 $x \in E$，皆有 $x \leqslant b$，2) 不存在 $b' < b$ 也具有性质 1)，那么就称 b 为集合 E 的上确界. 记之为

$$b = \sup E.$$

我们将证明由戴德金准则可以推出下面的定理：每一非空有上界的集合 E 都有上确界. 实际上，我们将所有的实数按照下面的方法分成两类 X 与 Y. 若有 $a \in E$ 及 $a \geqslant x$，则 $x \in X$；若对于任何 $a \in E$ 皆有 $a < y$，则 $y \in Y$. 不难看出，这是一个分割. 由戴德金准则，它有界 ξ，这个界就是集合 E 的上确界.

现在来证明由戴德金准则可以推出魏尔斯特拉斯准则. 命 $\{x_n\}$ 为递增有上界的实数序列，按刚才所证明的，它有上确界 ξ. 由上确界的定义可知，$x_n \leqslant \xi (n=1, 2, \cdots)$，且对于任何 $\varepsilon>0$，皆可以找到 n_0，使 $x_{n_0} > \xi - \varepsilon$. 由于序列 $\{x_n\}$ 的单调性，立刻得知对于所有的 $n > n_0$ 皆有 $\xi - \varepsilon < x_n \leqslant \xi$. 这就证明了 $\{x_n\}$ 以 ξ 为极限.

为了证明戴德金准则与魏尔斯特拉斯准则之间相反的关系，我们先注意，由魏尔斯特拉斯准则可以推出下面著名的准则。

阿基米德准则 对于任意实数 $a>0$ 与 b，皆有自然数 n，使 $na>b$。

这一准则说明了对于任何实数 b，序列 $\left\{\frac{b}{n}\right\}(n=1, 2, \cdots)$ 收敛于零。

如果魏尔斯特拉斯准则成立，而阿基米德准则不成立，那么就有这样的 $a>0$，使序列 $x_n=na$ 有界。因为它是递增的，由魏尔斯特拉斯准则，知它有极限 ξ。因此区间 $\left[\xi-\frac{a}{2}, \xi\right]$ 包含有序列的某个点 $x_n=na$。但是 $x_{n+1}=(n+1)a>\xi$，这与 ξ 是 $\{x_n\}$ 的上确界相矛盾。

由魏尔斯特拉斯准则可推出戴德金准则。将所有的实数分成两个无公共元素的集合 X 与 Y，对于任何 $x\in X$，$y\in Y$ 都有 $x<y$。我们来证明分割具有唯一的界 ξ。命 m 为整数，n 为自然数。记 x_n 是形状为 $\frac{m}{2^n}\in X$ 的数中最大的一个，此处 $x_n+\frac{1}{2^n}\in Y$。因为形状为 $\frac{m}{2^n}$ 的数包含在形状为 $\frac{m}{2^{n+1}}$ 的数的集合之中，所以 $x_n\leqslant x_{n+1}$。除此而外，序列 $\{x_n\}$ 是有界的 $\left(\text{例如不超过数 } x_1+\frac{1}{2}\right)$。所以由魏尔斯特拉斯准则，它有某个极限 ξ。我们来证明 ξ 就是我们的分割的界。事实上，假若 $x<\xi$，那么 $x\in X$。若 $y>\xi$，由阿基米德准则，可知存在 n，使 $\frac{1}{2^n}<y-\xi=a$，但 $x_n<\xi$，$x_n+\frac{1}{2^n}\in Y$，而 $y=\xi+a>x_n+\frac{1}{2^n}$，因此 $y\in Y$。

同样可以建立康托尔准则与柯西准则的等价性。然而，有时例如由柯西准则的成立，并不能推出戴德金准则的成立。正确地说，这句话应该了解为下面的意义：存在一个良序域，对于它，柯西准则成立而戴德金准则不成立。但假若事先就已假定了阿基米德准则成立，那么所有这四个准则都是等价的。

连续统非可数无限 我们现在来证明，间隔 $0 \leqslant x \leqslant 1$ 中全体点所成的集合不是可数无限的. 假定间隔 $0 \leqslant x \leqslant 1$ 中全体点所成的集合是可数无限的，那么可以借助于自然数，将这个间隔中所有的点 x 标记成为

$$x_1,\ x_2,\ \cdots,\ x_n,\ \cdots \tag{1}$$

在间隔 $[0, 1]$ 中，选取长度小于 1 的间隔 σ_1，它不包含 x_1. 这样的间隔确实能找到. 然后在 σ_1 里找出长度不超过 $\frac{1}{2}$ 的间隔 σ_2，而 σ_2 不包含 x_1 与 x_2. 一般说来，当我们已经选好了 σ_{n-1} 之后，我们再在里面截取长度不超过 $\frac{1}{n}$ 的间隔 σ_n，而 σ_n 不包含点 x_1, x_2, $\cdots$, x_n. 用这样的方法建立了间隔的无限序列

$$\sigma_1,\ \sigma_2,\ \cdots,\ \sigma_n,\ \cdots,$$

序列中每个间隔都包含在它前面的那个间隔之中，而这些间隔的长度，当 n 趋近于无限时趋于零. 由康托尔准则（见第 14 页）可知，间隔 $[0, 1]$ 中有唯一的点 x，属于所有的间隔 σ_n. 因为按照我们的假定，$[0, 1]$ 中所有的点都已经写入序列(1)，所以所有的 σ_n 的公共点 x，必定与这个序列中的某个点 x_m 相一致. 但是按照 σ_m 的构造，它不包含 x_m，因此 $x \neq x_m$. 这样一来，我们得到一个矛盾的结果. 所以假定 $[0, 1]$ 中所有的点所成的集合是可数无限集合是不正确的. 因此间隔 $[0, 1]$ 中所有的点不是可数无限的. 命题证完.

这一定理说明了存在不同的无限势，因此也就给第 8 页中第一个问题以正面的回答.

§ 4. 点　　集

在前一节里，我们已经遇到过以点为元素的集合，特别，我们考察了某个间隔全部点做成的集合，正方形 $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$ 内所有的点 (x, y) 做成的集合. 现在我们将要更加仔细地来研究这些集合的性质.

元素为点的集合称为点集，因此有直线上的点集，平面上的点集，任意空间中的点集．为了简单起见，我们仅仅考虑直线上的点集．

直线上的点与实数之间存在着紧密的联系．每一个实数都可以在直线上找到它所对应的点，且其逆也真．所以谈到点集，总是把它看作是由实数作成的集合——数轴上的集合．反之，我们也常常由集合内所有点的坐标来给出直线上的点集．

点集(正确地说是直线上的点集)具有一系列特殊的性质，使之与一般的集合有所区别，从而点集本身的理论也是数学的一个分支．首先两点之间的距离是有意义的，其次，在直线上的点之间可以建立次序的关系(左或右)，也就是说直线上的点是良序集合．最后，我们已指出，对于直线来说，康托尔准则是正确的．这一性质直接刻划了直线的稠密性．

我们来引进直线上最简单的集合的记号．

间隔 $[a, b]$ 表示坐标满足不等式 $a\leqslant x\leqslant b$ 的点的集合．

区间 (a, b) 表示坐标满足条件 $a<x<b$ 的点的集合．

半区间 $[a, b)$ 与 $(a, b]$，分别表示适合条件 $a\leqslant x<b$ 与 $a<x\leqslant b$ 的点的集合．

区间或半区间可能是非正规的．例如 $(-\infty, \infty)$ 表示直线上点的全体．又例如 $(-\infty, b]$ 表示满足 $x\leqslant b$ 的点的全体做成的集合．

我们首先来研究集合在直线上的位置的种种可能性．

有界的与无界的集合　直线上的点集 E，或者是由与原点相距不超过某一正常数的点所做成的，或者它具有与原点相距超过任意给定的大的常数的点．在前一种情况，我们称 E 为有界的集合；在后一种情况，则称 E 为无界的集合．例如间隔 $[0, 1]$ 中全体点所成的集合为有界集合，而全体整数所成的集合则为无界集合．

不难看出，如果 a 是直线上的一个固定的点，那么 E 有界的充分且必要的条件为，对于任何 $x\in E$，x 与 a 的距离都不超过某一正常数．

有上界与有下界的集合 命 E 为直线上的点集，若在直线上存在一个点 A，使 E 的任何点 x 都位于 A 的左方，则称集合 E 有上界．类似地，如果直线上有一点 a，使 E 的任何点 x 都位于 a 的右方，则称集合 E 有下界．这样一来，直线上有正坐标的全体点所成的集合有下界，而有负坐标的全体点所成的集合，则有上界．

若直线上的集合 E 有上界与下界，则称 E 是有界的．那么显然，上面引入的关于有界集合的定义是与此等价的．虽然这两种定义是如此相似，但在它们之间还是存在着本质的差别，第一个定义基于点之间有一个确定的距离，第二个定义则基于点集是良序集．

如果集合位于某间隔 $[a, b]$ 之中，则称此集合为有界集合，这也是定义有界集合的一种方法．

集合的上确界与下确界 命 E 是有上界的集合，那么在直线上存在点 A，在其右方没有集合 E 的任何点．由康托尔准则，可知在诸 A 之中，有一个位于最右端．这点称为集合 E 的上确界．同样可以定义点集的下确界．

如果集合 E 本身有一点位于所有点的右方，那么这一点就是 E 的上确界．但是我们不妨假定集合 E 本身没有位于最右端的点，例如坐标为

$$\frac{0}{1}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \cdots$$

的点集有上确界，但没有位于最右方的点．在这种情形之下，集合 E 的上确界 a 不属于 E，但与 a 相距任意近的地方，都有 E 的点．在上例中，$a=1$．

直线上任意点对于点集的位置 命 E 为点集而 x 是直线上的任意点．我们来研究集合 E 与点 x 的各种可能的位置．有下列各种可能情形：

1．点 x 及与 x 充分接近的点都不属于集合 E．

2．点 x 不属于 E，但不管与 x 距离多近，都有集合 E 的另外的点．

3. 点 x 属于 E, 但所有与 x 足够接近的点, 都不属于 E.

4. 点 x 属于 E, 而不论距 x 多近, 都有集合 E 的以外的点.

在情形 1, 称 x 是集合 E 以外的点; 在情形 3, 称 x 为集合 E 的孤立点; 在情形 2 与 4, 则称 x 为集合 E 的极限点.

因此, 如果 $x\overline{\in}E$, 则点 x 或者是集合 E 以外的点, 或者是它的极限点. 但如果 $x\in E$, 则它或者是 E 的孤立点, 或者是 E 的极限点.

极限点可以属于集合 E, 也可以不属于集合 E. 极限点的特征是无论距它多么近, 都有集合 E 的点, 也就是说, 如果任意包含 x 的区间 δ 都包含无限多个属于集合 E 的点, 那么 x 就是点集 E 的极限点. 极限点的概念是点集理论最重要的概念之一.

如果点 x 及与 x 足够接近的点都属于集合 E, 那么称 x 为集合 E 的内点. 若点 x 既非 E 以外的点, 又非 E 的内点, 则称为点集 E 的界点.

我们试举若干个例子来说明所有这些概念.

例 1 命集合 E_1 是由坐标为

$$1,\ \frac{1}{2},\ \frac{1}{3},\ \cdots,\ \frac{1}{n},\ \cdots$$

的点组成的. 那么这个集合的每一个点都是它的孤立点. 0 是集合 E_1 的极限点(不属于此集合), 而直线上其余的点都是 E_1 以外的点.

例 2 命 E_2 是由 $[0, 1]$ 中所有有理数所成的集合. 这个集合没有孤立点. 间隔 $[0, 1]$ 中所有的点都是 E_2 的极限点, 而直线上剩下来的点则都是 E_2 以外的点. 显然在 E_2 的极限点之中, 有些属于 E_2, 有些则不属于 E_2.

例 3 命 E_3 为间隔 $[0, 1]$ 中所有的点所成的集合. 正如前例一样, 集合 E_3 没有孤立点. $[0, 1]$ 中所有的点都是 E_3 的极限点. 但和前例不同的是, E_3 所有的极限点都属于 E_3.

例 4 命 E_4 是由直线上以整数为坐标的点所成的集合. E_4 的每一个点都是孤立点. E_4 没有极限点.

我们注意，对于例 3，区间 (0, 1) 中每一点都是 E_3 的内点。而对于例 2，则间隔 [0, 1] 中的每一点都是 E_2 的界点。

由以上所举的例子可以看出，直线上的无限点集可以有孤立点(例如 E_1, E_4)，也可以没有(例如 E_2, E_3)孤立点。同样，它可以有内点(例如 E_3)，也可以没有内点(例如 E_1, E_2, E_4)。除例 4 中的集合 E_4 没有一个极限点以外，其余诸例中的集合都有极限点。下面即将证明的非常重要的定理指出，这个现象和集合 E_4 的无界性有着密切的关系。

波尔察诺-魏尔斯特拉斯定理 任何直线上的有界无限点集至少有一个极限点。

证 命 E 表示直线上的有界无限点集。因为 E 有界，所以它一定位于某个间隔 $[a, b]$ 之中。把这个间隔平分一下。因为集合 E 是无限点集，所以这两个间隔中至少有一个包含无限多个 E 的点。我们记这个间隔为 σ_1 (如果 $[a, b]$ 的两个小间隔都含有无限多个点集 E 的点，那么 σ_1 就取作例如左边的那个间隔)。然后，我们再把 σ_1 分成长度相等的两个间隔，在这两个间隔之中，至少有一个包含无限多个点集 E 的点。我们记这个间隔为 σ_2。我们不断地继续这个手续，即把间隔平分为两个等长的小间隔，而取那个包含无限多个点集 E 的点的那个小间隔。这样一来，我们获得了间隔的序列 $\sigma_1, \sigma_2, \cdots, \sigma_n, \cdots$。这个序列具有这样的性质：每一间隔 σ_{n+1} 包含在它前面的那个间隔 σ_n 之中；每一间隔都包含无限多个点集 E 的点；间隔 σ_n 的长度趋于零。前两个性质，由序列的构造立刻得知。我们注意，若 $[a, b]$ 的长度为 l，则间隔 σ_n 的长度为 $l/2^n$。由此得出最后一个性质的证明。由康托尔准则可知存在唯一的点 x，属于所有的间隔 σ_n。我们现在来证明 x 是点集 E 的极限点。为此只要证明，若 δ 是任一含有点 x 的区间，则 δ 含有无限多个点集 E 的点。因为每一 σ_n 包有点 x，而 σ_n 的长度趋于 0，所以当 n 充分大时，间隔 σ_n 就整个包含在 δ 之中。由于 σ_n 包有无限多个点集 E 的点，因此点 x 是点集 E 的极限点。定理证完。

习题 证明：如果集合 E 有上界，但没有位于最右端的点，那么它的上确界必为 E 的极限点(但不属于 E).

闭集与开集 点集理论的基本问题之一是研究各种类型的点集的性质. 我们将向读者介绍这个理论的两个例子. 在此我们要研究所谓闭集与开集.

若一个集合含有它所有的极限点，就称这个集合为闭集. 没有一个极限点的集合，也被称为闭集. 除了自己的极限点之外，闭集也可以包含孤立点. 如果集合的每一个点都是它的内点，则称此集合为开集.

我们来介绍一些闭集与开集的例子. 每一个间隔 $[a, b]$ 都是闭集. 每一个区间 (a, b) 都是开集. 非正则的半区间 $(-\infty, b]$ 与 $[a, \infty)$ 都是闭集. 而非正则的区间 $(-\infty, b)$ 与 (a, ∞) 都是开集. 直线上所有点的集合既是闭集又是开集. 常常把空集也看作是既闭又开的集合. 直线上任一有限多个点的集合都没有极限点，所以是闭集. 由下面的点

$$0,\ 1,\ \frac{1}{2},\ \frac{1}{3},\ \frac{1}{4},\ \cdots,\ \frac{1}{n},\ \cdots$$

做成的集合是闭的；这一集合有唯一的属于自己的极限点 $x=0$.

我们的问题在于弄清楚任意闭集与开集的构造是怎样的. 为此我们引进一系列有用的事实，但都不加以证明了.

1. 任意多个闭集的交集仍是闭集.

2. 任意多个开集的和集仍为开集.

3. 若闭集有上界，则它以自己的点为上确界. 同样，若闭集有下界，则它以自己的点为下确界.

命 E 为直线上任意点的集合. 直线上不属于集合 E 的点的全体，称为集合 E 的补集，示之以 CE. 显然，若 x 为集合 E 以外的点，则 x 是集合 CE 的内点. 其逆也真.

4. 若集合 F 是闭的，则其补集 CF 是开集. 其逆也真.

命题 4 说明了在闭集与开集之间有着非常紧密的联系：一个是另一个的补集. 正因为如此，我们只要研究闭集或开集两者之一

即可．一种类型的集合的性质熟悉后，另一种类型的集合的性质跟着也就清楚了．例如，每一开集都可以由直线上去掉若干个闭集而得到．

我们开始研究闭集的性质．先引进一个定义．命 F 为闭集．若区间 (a, b) 中任意点都不属于集合 F，但 a 与 b 属于集合 F，那么就称具有这样性质的区间 (a, b) 为集合 F 的邻接区间．如果 a（或 b）属于集合 F，而 F 与非正规的区间 (a, ∞)（或 $(-\infty, b)$）无公共交点．我们也称这种非正规的区间为集合 F 的邻接区间．我们来证明．若 x 不属于闭集 F，则它一定属于集合 F 的某一邻接区间．

我们以 F_x 表示集合 F 位于 x 右方的部分．因为点 x 本身不属于集合 F，所以 F_x 可以表示为交集的形状

$$F_x = F \cdot [x, \infty).$$

因为集合 F 与 $[x, \infty)$ 都是闭集，因此由命题 1 得知集合 F_x 也是闭的．若集合 F_x 为空集，那么整个半区间 $[x, \infty)$ 中都没有集合 F 的点．现在我们假定集合 F_x 是非空的．因为 F_x 整个位于半区间 $[x, \infty)$ 中，所以，它有下界，我们以 b 表示它的下确界．由命题 3 可知，$b \in F_x$，所以 $b \in F$．进而言之，因为 b 是集合 F_x 的下确界，所以半区间 $[x, b)$ 位于 b 之左方，且不含集合 F_x 的点，所以不含有集合 F 的点．因此我们建立了半区间 $[x, b)$，它不包含集合 F 的点，而 b 属于集合 F 或 $b=\infty$．类似地可以建立 $(a, x]$，它也不包含集合 F 的点，且 $a=-\infty$ 或 $a \in F$．现在已经证明了区间 (a, b) 包含点 x，并且是集合 F 的邻接区间．不难看出，假若 (a_1, b_1) 与 (a_2, b_2) 是集合 F 的两个邻接区间，那么这两个区间或者重叠，或者无交点．

由上述立刻推知，直线上每一个闭集 F，都可以由直线上去掉若干个区间而得到，即去掉集合 F 的邻接区间而得到．因为每一个区间中至少包含一个有理点，而直线上所有的有理点是可数无限集合，由此可知，所有的邻接区间的个数不多于可数无限多个．故得最后的结论：每一直线上的闭集，可以由直线上去掉不超过可

数无限多个不相交的区间而得到.

由命题 4 立即推知，每一直线上的开集都可以表为不超过可数无限多个不相交的区间之和. 由命题 1 与 2 同样也得知，如上法构成的集合，的确是闭(开)的.

由下面的例子可以看出，闭集可以有非常复杂的结构.

康托尔完全集合 我们建立一个特殊的闭集，它有许多奇妙的性质. 首先在直线上移去非正规的区间 $(-\infty, 0)$ 与 $(1, \infty)$，于是我们得到间隔 $[0, 1]$. 我们再移去这个间隔中间的三分之一的那个区间 $\left(\frac{1}{3}, \frac{2}{3}\right)$. 然后，在每个留下来的间隔 $\left[0, \frac{1}{3}\right]$ 与 $\left[\frac{2}{3}, 1\right]$ 中，移去其中间的三分之一的那个区间. 这种移去剩下来的间隔的中间的三分之一那个区间的手续，可以不断地进行下去. 把所有这些区间移去后，在直线上剩下来的点的集合，称为康托尔完全集合. 我们以字母 P 示之.

我们来研究这个集合的某些性质. 因为集合 P 是由直线上不断移去若干不相交的区间而得到的，所以集合 P 为闭集. 又因为在每次被移去的区间的端点都属于 P，所以集合 P 是非空的.

如果闭集不包含孤立点，也就是说这个集合所有的点都是极限点，那么就称闭集 F 为完全集. 现在我们来证明，集合 P 是完全集. 实际上，如果某个点 a 是集合 P 的孤立点，那么它就是这个集合的某两个邻接区间的公共端点. 但是由集合 P 的构造可以看出，它的邻接区间并无公共端点.

集合 P 中不包含任何区间. 事实上，假若有某个区间 δ 整个都属于集合 P，那么，它一定属于构造 P 的过程的第 n 步所获得的间隔之一. 但这是不可能的. 此乃因为当 $n\to\infty$ 时每个间隔的长度都趋于零.

可以证明，P 具有连续统势. 特别由此推出康托尔完全集除了包含邻接区间的端点外，还包含一些其他的点. 此乃由于邻接区间仅仅构成可数无限集.

各种不同类型的点集常常出现在各个数学分支之中，对于研

究许多数学问题，熟悉这些性质是完全必要的．点集的理论对于数学分析及拓扑学有着根本性的重要意义．

我们试举若干古典分析分支内的点集的例子．命 $f(x)$ 是间隔 $[a, b]$ 上的连续函数．固定数 α，考虑适合 $f(x)\geqslant\alpha$ 的点 x 的集合．不难证明，这个集合是间隔 $[a, b]$ 上的闭集．同样可知适合 $f(x)>\alpha$ 的点 x 的集合，是区间 (a, b) 上的开集 G．若 $f_1(x)$, $f_2(x)$, $\cdots$, $f_n(x)$, $\cdots$ 是间隔 $[a, b]$ 上的连续函数序列，则使这个函数序列收敛的点 x 的集合，不会是任意的，而是属于一个完全确定的类型的点集．

研究点集构造的数学分支称为描述集合论．苏联数学家——鲁金及其学生亚历克山大洛夫、苏斯林、阔尔莫果洛夫、拉夫抡捷夫、诺维科夫、凯尔迪什、李雅普诺夫等等，在描述集合论的发展上，树立了巨大的功勋．

鲁金和他的学生的研究，阐明了在描述集合论与数理逻辑之间有着深刻的联系．描述集合论所引起的困难(特别是关于各种集合的势的定义问题)是属于逻辑性的困难．反之，数理逻辑的方法能更深刻地去洞察描述集合论方面的某些问题．

§5. 集合的测度

集合的测度概念是间隔长度的概念的进一步推广．在最简单的情形(也就是现在所仅仅考虑的情况)下，问题在于如何给出不仅是对于间隔而且是对于直线上更为复杂的点集的长度概念．

我们给间隔 $[0, 1]$ 以长度 1．那么显然，任意间隔 $[a, b]$ 的长度就等于 $b-a$．同样，如果有两个不相交的间隔 $[a_1, b_1]$ 与 $[a_2, b_2]$，那么由这两个间隔的点所构成的集合 E 的长度显然就是 $(b_1-a_1)+(b_2-a_2)$．然而如何得出直线上更为复杂的集合的长度，仍然是不清楚的．例如本章 §4 所讨论的康托尔集合 P 的长度等于什么？因此，直线上集合的长度概念需要有严格的数学意义．

集合的长度的定义问题，或者说是集合的测度问题，是非常重要的．因为它对于区间概念的推广有着本质的意义．集合的测度的概念在函数论的一些问题上常用到，同样在概率论、拓扑学、泛函分析等方面也用到．

以下我们来阐述集合的测度的定义．这是法国数学家勒贝格提出的．这一定义是他定义积分的基础．

开集与闭集的测度　首先我们来定义任意开集与任意闭集的测度．在§4已经阐明，直线上的每一个开集都是有限多个或可数无限多个两两不相交的区间的和．

构成开集的区间的长度之和称为开集的测度．

因此，如果

$$G=\sum(a_i,\ b_i)$$

而区间$(a_i,\ b_i)$是两两不相交的，那么G的测度就等于$\sum(b_i-a_i)$．我们一般用μE来记集合E的测度．因此

$$\mu G=\sum(b_i-a_i).$$

特别可知，一个区间的测度等于它的长度

$$\mu(a,\ b)=b-a.$$

每一个包含在间隔$[a,\ b]$中的闭集F，只要间隔$[a,\ b]$的端点属于F，都可以由$[a,\ b]$中去掉某个开集G而得到．因此，间隔$[a,\ b]$的长度与F的补集（关于$[a,\ b]$），即开集G的测度之差就称为闭集$F\subseteq[a,\ b]$的测度，此处$a\in F$，$b\in F$．

因此

$$\mu F=(b-a)-\mu G. \tag{2}$$

不难看出，依照这个定义，任意间隔的测度等于它的长度

$$\mu[a,\ b]=b-a.$$

而由有限多个点做成的集合的测度等于零．

测度的一般定义　为了给出比闭集与开集更为一般的集合的测度的定义，我们需要一个辅助概念．命E是某个间隔$[a,\ b]$上的集合．我们考虑集合E所有的可能的遮盖，即所有的包含集合E的可能的开集$V(E)$．每一个集合$V(E)$的测度都已经确定好

了．所有的集合 $V(E)$ 的测度的全体是某些正数的集合．这个数集有下界(例如数零)，因此有下确界．我们记之为 $\mu_e E$．数 $\mu_e E$ 称为集合 E 的外测度．

命 $\mu_e E$ 表示集合 E 的外测度，而 $\mu_e CE$ 表示它对于间隔 $[a, b]$ 的补集的外测度．

如果关系式

$$\mu_e E+\mu_e CE=b-a \tag{3}$$

成立，就称集合 E 是可测度的．而数 $\mu_e E$ 就称为它的测度：$\mu E=\mu_e E$；如果关系式(3)不满足，就称集合 E 是不可测度的．不可测的集合不具有测度．

我们注意

$$\mu_e E+\mu_e CE\geqslant b-a \tag{4}$$

总是成立的．

我们作些说明．简单的集合(例如区间与间隔)的长度具有许多值得注意的性质．现在列举其中重要的一些．

1．假若集合 E_1 与 E 都是可测的，而且 $E_1\subseteq E$，那么

$$\mu E_1\leqslant \mu E.$$

即集合 E 的一部分的测度不超过整个集合 E 的测度．

2．假若集合 E_1 与 E_2 都可测，那么集合 $E=E_1+E_2$ 也可测，而且

$$\mu(E_1+E_2)\leqslant \mu E_1+\mu E_2.$$

即集合的和的测度不超过集合的测度之和．

3．假若集合 $E_i(i=1, 2, \cdots)$ 都可测，而且两两不相交：$E_iE_j=\theta(i\neq j)$，则其和 $E=\sum E_i$ 也可测，且

$$\mu(\sum E_i)=\sum \mu E_i.$$

即有限多个或可数无限多个两两不相交的集合之和的测度等于这些集合的测度之和．

称测度的这一性质为它的完全可加性．

4．如果集合经过刚体运动，它的测度是不改变的．

可否期望，长度的基本性质对于更一般的测度来说，仍然保

存．假如我们给直线上任意点集以测度的话，可以完全严格地证明这是不可能的．所以根据上面所给出的定义，有可测的，即有测度的集合，也有不可测的，即没有测度的集合．由于可测集合是如此之多，以致于我们并不感到有任何本质的不方便．甚至连造出一个不可测的集合的例子，也还不是很容易的．

我们试举某些可测集合的例子．

例 1 康托尔完全集合 P 的测度（见 § 4）．P 是这样构造的：先从间隔 $[0, 1]$ 中去掉一个长度为 $\frac{1}{3}$ 的邻接区间，再去掉两个长度为 $\frac{1}{9}$ 的邻接区间，然后是四个长度为 $\frac{1}{27}$ 的邻接区间，如此等等．一般说来，在第 n 步时，去掉 2^{n-1} 个长度都是 $\frac{1}{3^n}$ 的邻接区间．于是，所有去掉的区间的长度之和等于

$$S=\frac{1}{3}+\frac{2}{9}+\frac{4}{27}+\cdots+\frac{2^{n-1}}{3^n}+\cdots.$$

这一级数是几何级数，首项等于 $\frac{1}{3}$ 而公比为 $\frac{2}{3}$．于是级数 S 的和等于

$$\frac{\frac{1}{3}}{1-\frac{2}{3}}=1.$$

因此，康托尔完全集合所有的邻接区间的长度之和等于 1．换言之，P 的补集，即开集 G 的测度为 1．因此集合 P 本身的测度为

$$\mu P=1-\mu G=1-1=0.$$

这个例子说明了集合可以具有连续统势，但是它的测度却是零．

例 2 间隔 $[0, 1]$ 中所有有理数 R 的集合的测度．首先我们证明 $\mu_e R=0$．在 § 2 中已经介绍过 R 是可数无限的．将 R 中的点排成序列

$$r_1, r_2, \cdots, r_n, \cdots.$$

给出 $\varepsilon>0$，作包含 r_n 的区间 δ_n，其长度为 $\frac{\varepsilon}{2^n}$。和 $\delta=\sum\delta_n$ 是开集，它遮住了 R。因为区间 δ_n 可以彼此相交，所以

$$\mu(\delta)=\mu(\sum\delta_n)\leqslant\sum\mu\delta_n=\sum\frac{\varepsilon}{2^n}=\varepsilon.$$

因为 ε 可以选取得任意小，所以 $\mu_e R=0$。

进而言之，由(3)得

$$\mu_e R+\mu_e CR\geqslant 1,$$

即 $\mu_e CR\geqslant 1$。因为 CR 包含在间隔 $[0,1]$ 之中，所以 $\mu_e CR\leqslant 1$。因此

$$\mu_e R+\mu_e CR=1.$$

故得

$$\mu R=0,\ \mu CR=1^{1)} \tag{5}$$

这一例子说明了集合可以在某一区间上处处稠密，但其测度却是零。

测度为零的集合在函数论的许多问题中都不起主要作用，而往往是可以忽略不计的。例如，函数 $f(x)$ 在黎曼意义下可以求积分的充要条件为它是有界的，且其不连续点的测度为零。像这样的例子还可以举出很多来。

可测函数 我们现在考虑集合的测度概念的最主要的应用之一，那就是用它来描绘函数论与数学分析中常常处理的一类函数。问题正确的提法是，在某个集合 E 上给出函数序列 $\{f_n(x)\}$，除了集合 E 上一个测度为零的集合 N 之外，在集合 E 的其他点都收敛，那么我们就称函数序列 $\{f_n(x)\}$ 几乎处处收敛。

怎样的函数可以由反复地运用几乎处处收敛的连续函数序列的趋限运算及代数运算得到呢？

为了回答这个问题，我们需要某些新的概念。

命函数 $f(x)$ 在集合 E 上定义，而 α 是任意实数。我们以

$$E[f(x)>\alpha]$$

表示 E 中适合 $f(x)>\alpha$ 的那些点。例如，假若函数 $f(x)$ 在间隔

1) 这一推理的本身就说明了，直线上的任一可数无限点集的测度皆为零。

$[0,1]$ 上定义，而且在这个间隔上 $f(x)=x$，那么当 $\alpha<0$ 时，$E[f(x)>\alpha]$ 为 $[0, 1]$；当 $0\leqslant\alpha<1$ 时为 $(\alpha,1]$；而当 $\alpha\geqslant1$ 时，则为空集.

函数 $f(x)$ 在某个集合 E 上定义. 如果 E 本身可测，且对于任何实数 α，集合 $E[f(x)>\alpha]$ 也可测，就称函数 $f(x)$ 是可测函数.

可以证明，在间隔上给出的任意连续函数是可测的. 但是，有许多不连续的函数，同样也是可测函数. 例如狄利克雷函数，即在间隔 $[0, 1]$ 的无理点上等于 1 而在其余的点上等于零的那个函数，即不连续而可测.

可测函数具有以下诸性质. 在此我们不给出它们的证明了.

1. 假若 $f(x)$ 与 $\varphi(x)$ 是在同样的集合 E 上定义的可测函数，那么函数

$$f+\varphi,\ f-\varphi,\ f\cdot\varphi \text{ 与 } \frac{f}{\varphi}$$

都是可测的(在最后一个关系中，须假定 $\varphi\neq0$).

这一性质说明了在可测函数上进行代数运算，重新得到可测函数.

2. 如果 $\{f_n(x)\}$ 是在集合 E 上定义的可测函数序列，几乎处处收敛于函数 $f(x)$，那么这个函数同样是可测的.

因此，几乎处处收敛的可测函数序列，经过趋限运算后，重新得到可测函数.

可测函数的这些性质是勒贝格确定的. 可测函数更深刻的研究是苏联数学家鲁金及叶果洛夫完成的. 特别是鲁金证明了间隔上的每一可测函数，只要改变测度任意小的集合上的数值，就变为连续函数.

鲁金的这一经典结果及以上所列举的可测函数的性质证明了可测函数就是本段开端所要问的那类函数. 可测函数对于积分论也有着巨大的意义，那就是积分的概念可以推广，使每一有界可测函数都是可以求积分的.关于这一点，在下一节将给出详细的阐述.

§6. 勒贝格积分

我们来讨论本章的中心问题，即勒贝格积分的定义及其性质.

为了了解这个积分的构造原理，我们来观察下面的例子. 设有大量不同价值的钱币，要求计算这些钱币的总值. 可以有两种不同的计算方法. 固然可以用任意依次累加钱的价值的方法，然而也可以用另外的方法来计算，就是用将钱分堆加的办法. 将同

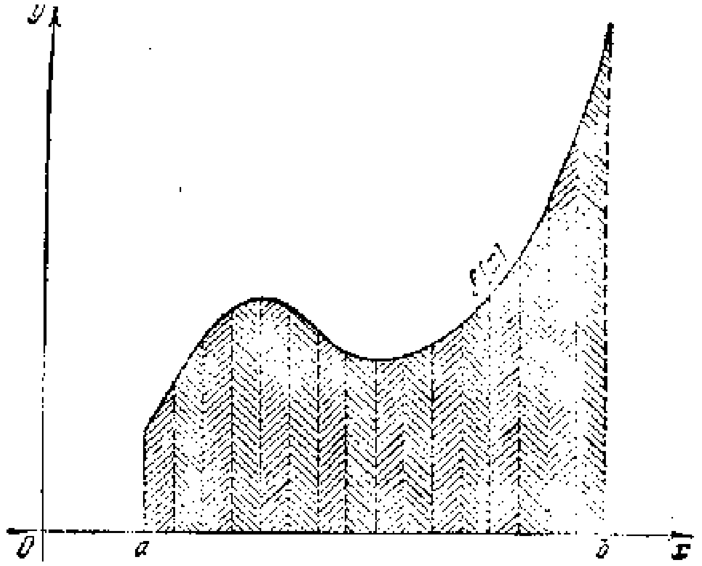


图2 a

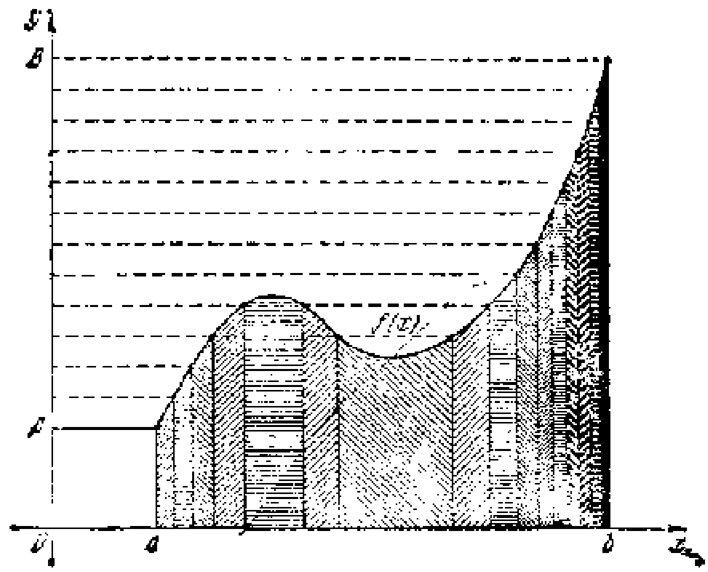


图2 b

样价值的钱币放在一堆，再将每一堆钱币的个数乘上这堆钱币的单价，然后将所得到的这些数值都加起来即可．第一种计算钱币的方法，对应于黎曼的积分过程，而第二种则对应于勒贝格积分过程．

现在我们由钱币转而讨论函数．我们可以说，黎曼积分的运算是将给定函数的定义区域（坐标横轴，见图 2 a）分小而产生的，而勒贝格积分的计算则是划分函数值的区域（坐标纵轴，见图 2 b）而产生的．后一种远在勒贝格之前，当人们计算带有振动性质的函数的积分时，已经用到过．但是勒贝格首先一般地发展了它，并且借助于测度论，给它以严格的论证．

我们来观察集合的测度与勒贝格积分之间的联系是怎样的．命 E 是某个间隔 $[a,b]$ 上的任意的可测集合．我们造出函数 $\varphi(x)$，当 x 属于 E 时，其值为 1；当 x 不属于 E 时，其值为零，也就是说

$$\varphi(x)=\begin{cases}1, & x\in E;\\ 0, & x\overline{\in} E.\end{cases}$$

函数 $\varphi(x)$ 常常称为集合 E 的特征函数．现在来考虑积分

$$I=\int_a^b \varphi(x)dx.$$

我们已经习惯于认为这个积分等于图形 D 的面积，此处 D 是由横坐标、直线 $x=a$、$x=b$ 及曲线 $y=\varphi(x)$ 所范围起来的（见第一卷第二章）．因为图形 D 的“高”当且仅当 $x\in E$ 时异于零，而其值为 1，所以（按照公式，面积等于长度乘上宽度），它的面积应该算作是集合 E 的长度（测度）．因此，I 应该等于集合 E 的测度

$$I=\mu E. \tag{6}$$

这也就是关于函数 $\varphi(x)$ 的勒贝格积分的定义．

读者应该确切理解，等式(6)只是将 $\int_a^b \varphi(x)dx$ 看成勒贝格积分的积分定义．有时会发生这样的情况，I 在第二章（第一卷）所考虑的意义之下是不存在的，即积分和的极限是不存在的．但是如果后者存在，那么把 I 看作勒贝格积分时也存在，且等于 μE．

我们来计算狄利克雷函数 $\Phi(x)$ 的积分作为一个例子．$\Phi(x)$

在间隔[0, 1]的无理点上取值1，而在有理点上取值0．因为按照(5)，间隔[0, 1]上的无理点的集合的测度为1，故勒贝格积分

$$\int_0^1 \Phi(x)dx$$

等于1．不难证明，这一函数的黎曼积分是不存在的．

一个辅助命题 现在命$f(x)$是间隔$[a, b]$上的任意的有界可测函数．我们来证明每一个这样的函数都可以无限接近地表示为集合的特征函数的线性组合的形式．为了证明这一点，在函数值的下确界A与上确界B之间，添入分点$y_0=A, y_1, \cdots, y_n=B$，使每一小间隔的长度都不超过$\varepsilon$，此处$\varepsilon$为任意固定的小正数．进而言之，如果$x\in[a, b]$，

$$y_i\leqslant f(x)<y_{i+1} \quad (i=0,1,\cdots,n-1),$$

则在这些点上，置

$$\varphi(x)=y_i,$$

而在满足

$$f(x)=y_n=B$$

的点x上置

$$\varphi(x)=y_n.$$

函数$\varphi(x)$的结构见图3．

由于函数$\varphi(x)$的构造，在间隔$[a, b]$的任意点上皆满足

$$|f(x)-\varphi(x)|<\varepsilon.$$

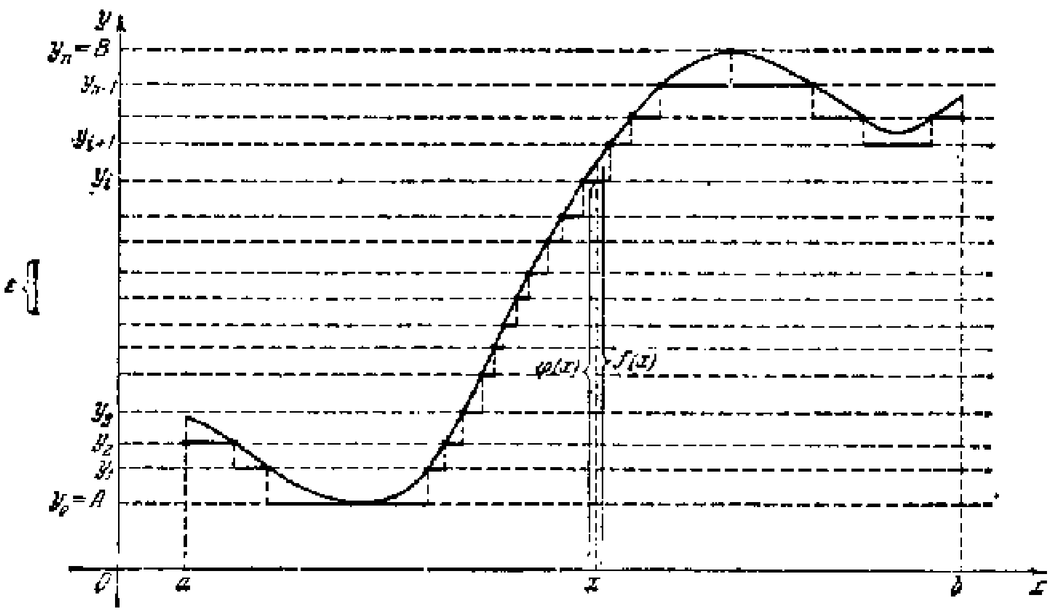


图 3

除此而外，因为函数 $\varphi(x)$ 仅取有限多个数值 $y_0, y_1, \cdots, y_n$，所以，它可以写为形状

$$\varphi(x)=y_0\cdot\varphi_0(x)+y_1\cdot\varphi_1(x)+\cdots+y_n\cdot\varphi_n(x), \tag{7}$$

此处 $\varphi_i(x)$ 是满足 $\varphi(x)=y_i$，即 $y_i\leqslant f(x)<y_{i+1}$ 的点的集合的特征函数(对于每一个点 $x\in[a, b]$，(7) 式右端仅有一个被加数不是零!)．这样一来，我们的命题就证完了．

勒贝格积分的定义 我们来谈谈任意有界可测函数的勒贝格积分的定义．因为函数 $\varphi(x)$ 与 $f(x)$ 的相差非常小，所以 $\varphi(x)$ 的积分可以当作 $f(x)$ 的积分的近似值．由于函数 $\varphi_i(x)$ 是集合的特征函数及一些普通的积分运算的法则，我们得到

$$\begin{aligned}\int_a^b \varphi(x)\,dx&=\int_a^b\{y_0\varphi_0(x)+y_1\varphi_1(x)+\cdots+y_n\varphi_n(x)\}\,dx\\&=y_0\int_a^b\varphi_0(x)\,dx+y_1\int_a^b\varphi_1(x)\,dx+\cdots+y_n\int_a^b\varphi_n(x)\,dx\\&=y_0\mu e_0+y_1\mu e_1+\cdots+y_n\mu e_n,\end{aligned}$$

此处 μe_i 是满足

$$y_i\leqslant f(x)<y_{i+1}$$

的那些点 x 的集合 e_i 的测度．所以，"勒贝格积分和"

$$S=y_0\mu e_0+y_1\mu e_1+\cdots+y_n\mu e_n$$

是函数 $f(x)$ 的勒贝格积分的近似值．因此当

$$\max|y_{i+1}-y_i|\to 0$$

时，函数 $\varphi(x)$ 一致收敛于其对应的 $f(x)$，勒贝格积分和 S 的极限就定义作勒贝格积分．

可以证明，对于任意的有界可测函数，勒贝格积分和是有极限的，即任意有界可测函数是勒贝格可积函数．勒贝格积分的定义还可以推广到某些非有界的可测函数类上去．但是，我们不准备在此讨论了．

勒贝格积分的性质 勒贝格积分具有与通常积分一样良好的性质．例如，函数和的积分等于函数的积分的和，常数因子可以移出积分记号等等．但是，勒贝格积分还有一个通常积分不具备的

非常重要的性质：如果对于所有的 n 及 $[a, b]$ 中的 x，有常数 K 使所有的可测函数 $f_n(x)$ 皆满足

$$|f_n(x)|<K,$$

且函数序列 $\{f_n(x)\}$ 又几乎处处收敛于 $f(x)$，那么

$$\int_a^b f_n(x)dx \to \int_a^b f(x)dx.$$

换言之，勒贝格积分允许在限制不强的条件下进行积分号内趋限。正由于勒贝格积分有这一性质，所以在许多研究中，它常常成为不可避免的与很方便的工具。特别在三角级数论、函数空间论及其他许多数学分支中，勒贝格积分都是完全必需的。

我们举一个例子。命 $f(x)$ 是周期为 2π 的周期函数，而

$$\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos nx+b_n\sin nx)$$

为它的富里埃级数。例如，假定 $f(x)$ 是连续的，那么，不难证明

$$\frac{1}{\pi}\int_0^{2\pi} f^2(x)\,dx=\frac{a_0^2}{2}+\sum_{n=1}^{\infty}(a_n^2+b_n^2). \tag{8}$$

这个恒等式通常称为巴斯娃等式。我们考虑这样的问题：怎样的周期函数，才能使巴斯娃等式(8)成立？这一问题的回答是：巴斯娃等式(8)当且仅当函数 $f(x)$ 为间隔 $[0, 2\pi]$ 上的可测函数，而函数 $f^2(x)$ 为这一间隔上的勒贝格可积函数时才成立。

文　献

А. Лебег, Об измерении величин. ГОНТИ, 1938.

А. И. Маркушович, Действительные числа и основные принципы теория пределов. Учпедгиз, 1948.

П. С. 亚历山大洛夫，集与函数的泛论初阶，商务印书馆，1954（共上、下两册）.

П. С. Александров и А. Н. Колмогоров, Введение в теорию функций действительного переменного. Изд. 3-е, ГОНТИ, 1938.

И. П. 那汤松，实变函数论，高等教育出版社，1955（共上、下两册）.

А. Лебег, Интегрирование и отыскание примитивных. ГТТИ, 1934.

Ф. Хаусдорф, Теория множеств. ГОНТИ, 1937.

王　元　译

赵民义　校

第十六章　线性代数

§1. 线性代数的对象和它的工具

线性函数和矩阵　在一个变数的函数中，线性函数 $l(x)=ax+b$ 是最简单的函数. 大家都知道，它的图象是最简单的线——直线.

同时，线性函数还是最重要的函数之一. 这是因为任意一条"光滑"曲线的一小段都很像直线，而且曲线的这一小段愈短，它就愈接近于直线. 用函数论的语言来说，这意思就是，任意"光滑"（连续可微分的）函数，当独立变数改变很小时，接近于线性函数. 线性函数可以刻划作它的增量与独立变数的增量成比例的函数. 实际上，$\Delta l(x)=l(x_0+\Delta x)-l(x_0)=a(x_0+\Delta x)+b-(ax_0+b)=a\Delta x$. 反之，如果 $\Delta l(x)=a\Delta x$，则 $l(x)-l(x_0)=a(x-x_0)$ 及 $l(x)=ax+l(x_0)-ax_0=ax+b$，而 $b=l(x_0)-ax_0$. 从微分学我们知道，在任意可微函数的增量中很自然地分出来一主要部分，即所谓的函数的微分，它与独立变数的增量成比例，而且函数的增量与其微分相差只是独立变数的增量的高阶无穷小. 这样一来，可微函数在独立变数改变很小时，实际上接近于线性函数. 所差的只是一个高阶无穷小.

对于多元函数也有类似的情形.

形状如 $a_1x_1+a_2x_2+\cdots+a_nx_n+b$ 的函数称为多元线性函数. 如果 $b=0$，这个线性函数就称为齐次函数. 以下两个性质是多元线性函数的特征：

1. 假定只让一个独立变数获得某个增量而让其余的变数不改变，这样算得的线性函数的增量与这个独立变数的增量成比例.

2. 假定所有独立变数都得到某些增量，这样算得的线性函数

的增量等于单独改变每一个变数所得到的增量的代数和。

多元线性函数在多元函数中起着单元线性函数在单元函数中同样的作用. 这就是说, 任意"光滑"函数(即对于每一个变数都有连续偏微商的函数)当独立变数改变很小时, 接近于某个线性函数. 实际上, 这样一个函数 $w=f(x_1, x_2, \cdots, x_n)$ 的增量, 除了差一个高阶无穷小之外, 等于全微分 $\frac{\partial f}{\partial x_1}dx_1+\frac{\partial f}{\partial x_2}dx_2+\cdots+\frac{\partial f}{\partial x_n}dx_n$. 它是独立变数的增量 $dx_1, dx_2, \cdots, dx_n$ 的线性齐次函数. 由此推出, 函数 w 本身等于它的开始值与增量的和, 它在独立变数改变很小时, 除了差一个高阶无穷小之外, 可以表作独立变数的增量的线性非齐次函数.

由于研究关联着多个因素的量所引起的问题, 解决起来需要考察多元函数. 如果所研究的关联性是线性的, 那么这个问题就称为线性问题. 基于以上所指出的线性函数的性质, 以下诸特性可以说明线性问题.

1. 比例性 每一个独立因素作用的结果都与它的大小成比例.

2. 无关性 作用的总结果等于个别因素的作用的结果之和.

任意"光滑"函数, 当独立变数改变很小时, 在它的第一级逼近的范围内可以用线性函数来代替这件事实是一个一般性原则的反映. 这个原则是, 在研究依赖于某些因素的作用的量时, 在作用很小的情况下, 关于该量变动的任何问题, 在它第一级逼近的范围内可以看作是线性问题, 即具有无关性与比例性的问题. 时常发现, 这样的看法得到满足实用要求的结果(古典弹性力学、微小振动的理论等等).

所研究的物理量本身常常可用一些数量来描写(利用它在坐标轴上的三个分量来描写, 弹性体在已知点的应变状态用所谓应变张量的六个分量来描写等等). 因此有必要同时研究一些多元函数, 而在第一级逼近范围内有必要同时研究一些线性函数.

一元线性函数在性质上相当简单, 以致不需要任何特别研究.

对于多元线性函数，事情却是另外一个样子，这时多个变数的存在引起了一些特殊的性质。如果从一些变数 $x_1, x_2, \cdots, x_n$ 的一个函数进而讨论这些变数的多个函数 $y_1, y_2, \cdots, y_m$ 的集合，事情就更为复杂。这时作为"第一级逼近"就出现了下面线性函数的集合：

$$
\begin{aligned}
y_1 &= a_{11}x_1 + \cdots + a_{1n}x_n + b_1, \\
y_2 &= a_{21}x_1 + \cdots + a_{2n}x_n + b_2, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
y_m &= a_{m1}x_1 + \cdots + a_{mn}x_n + b_m.
\end{aligned}
$$

线性函数的集合也是相当复杂的数学对象，它的研究是富有兴趣与不显然的内容。

关于线性函数和它们的集合的理论，首先是代数里所谓线性代数这一分支的对象。

历史上线性代数的第一个问题是关于解线性方程组

$$
\begin{aligned}
a_{11}x_1 + \cdots + a_{1n}x_n &= b_1, \\
a_{21}x_1 + \cdots + a_{2n}x_n &= b_2, \\
\cdots\cdots\cdots\cdots&\cdots\cdots \\
a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m
\end{aligned}
$$

的问题。

在中学初等代数课程里讨论过这个问题的最简单情形。对于较大的 n，寻求方程组尽可能最简单的而且计算最不复杂的数值解的方法，直到现在还受到很多学者的注意，因为方程组的数值解法是很多计算与研究中的重要组成部分。

线性齐次函数也称为**线性型**。一组已知的线性型

$$
\begin{aligned}
y_1 &= a_{11}x_1 + \cdots + a_{1n}x_n, \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
y_m &= a_{m1}x_1 + \cdots + a_{mn}x_n
\end{aligned}
$$

由它的系数所决定，因为这样一组线性型的性质只依赖于系数的数值，而变数的名字没有实质意义。

例如，线性型组

$$\begin{matrix} 3x_1+x_2-x_3, & & 3t_1+t_2-t_3, \\ 2x_1+x_2+3x_3, & \text{和} & 2t_1+t_2+t_3, \\ x_1-x_2-x_3 & & t_1-t_2-t_3 \end{matrix}$$

当然有同样的性质，因而可以看作没有本质的区别.

自然可以把线性型组的系数的全体排成一个长方形的表：

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

这种表称为矩阵. 数 a_{ij} 称为矩阵的元素. 从线性代数的对象本身必然产生研究矩阵的必要性.

矩阵的重要而特殊的情形有：由一列组成的矩阵，简称为列；由一行组成的矩阵，简称为行；以及方阵，即行数等于列数的矩阵. 方阵的行数(或列数)称为它的阶数. 由一个数 a 所组成的"矩阵"(a)就看作是这个数.

由于线性型组之间的简单运算，自然地定义了矩阵之间的简单运算.

设给了两组线性型

$$\begin{aligned} y_1 &= a_{11}x_1+\cdots+a_{1n}x_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m &= a_{m1}x_1+\cdots+a_{mn}x_n \end{aligned}$$

和

$$\begin{aligned} z_1 &= b_{11}x_1+\cdots+b_{1n}x_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ z_m &= b_{m1}x_1+\cdots+b_{mn}x_n. \end{aligned}$$

把这两组线性型分别加起来得到

$$\begin{aligned} y_1+z_1 &= (a_{11}+b_{11})x_1+\cdots+(a_{1n}+b_{1n})x_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m+z_m &= (a_{m1}+b_{m1})x_1+\cdots+(a_{mn}+b_{mn})x_n. \end{aligned}$$

由这组型所得到的矩阵

$$\begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ a_{m1}+b_{m1} & \cdots & a_{mn}+b_{mn} \end{pmatrix}$$

自然看作是矩阵

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots\cdots & \cdots & \cdots\cdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \text{和} \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \cdots\cdots & \cdots & \cdots\cdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}$$

的和.

类似地，矩阵$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots\cdots & \cdots & \cdots\cdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$乘以数 c 定义为线性型组 cy_1, cy_2, …, cy_m 的系数所组成的矩阵，其中 y_1, y_2, …, y_m 是线性型，它们的系数组成矩阵$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots\cdots & \cdots & \cdots\cdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$.

从这个定义显然有

$$c\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots\cdots & \cdots & \cdots\cdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} ca_{11} & \cdots & ca_{1n} \\ \cdots\cdots & \cdots & \cdots\cdots \\ ca_{m1} & \cdots & ca_{mn} \end{pmatrix}.$$

最后，矩阵与矩阵的乘法运算如下定义. 设

$$\left.\begin{aligned} z_1 &= a_{11}y_1+\cdots+a_{1m}y_m, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ z_k &= a_{k1}y_1+\cdots+a_{km}y_m \end{aligned}\right\} \tag{1}$$

而

$$\begin{aligned} y_1 &= b_{11}x_1+\cdots+b_{1n}x_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m &= b_{m1}x_1+\cdots+b_{mn}x_n. \end{aligned}$$

将 y_1, y_2, …, y_m 通过 x_1, x_2, …, x_n 的表达式代入(1)中得到 z_1, z_2, …, z_k 通过 x_1, x_2, …, x_n 表成的线性型

$$\begin{aligned} z_1 &= c_{11}x_1+\cdots+c_{1n}x_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}$$

$$z_k = c_{k1}x_1 + \cdots + c_{kn}x_n.$$

这组线性型的系数矩阵

$$\begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \cdots & \cdots & \cdots \\ c_{k1} & \cdots & c_{kn} \end{pmatrix}$$

称为矩阵 $\begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \cdots & \cdots & \cdots \\ a_{k1} & \cdots & a_{km} \end{pmatrix}$ 和 $\begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \cdots & \cdots & \cdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}$

的乘积，记作 $\begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \cdots & \cdots & \cdots \\ a_{k1} & \cdots & a_{km} \end{pmatrix}\begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \cdots & \cdots & \cdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}.$

利用两个矩阵的元素不难算出两个矩阵的乘积的元素．设 c_{ij} 是用 $x_1, \cdots, x_n$ 表 z_i 时 x_j 的系数．因为 $z_i = a_{i1}y_1 + \cdots + a_{im}y_m$，而

$$\begin{aligned} y_1 &= \cdots + b_{1j}x_j + \cdots, \\ &\cdots\cdots\cdots\cdots \\ y_m &= \cdots + b_{mj}x_j + \cdots, \end{aligned}$$

所以 $$z_i = \cdots + (a_{i1}b_{1j} + \cdots + a_{im}b_{mj})x_j + \cdots.$$

由此推出 $$c_{ij} = a_{i1}b_{1j} + \cdots + a_{im}b_{mj}.$$

因此，两个矩阵的乘积中第 i 行第 j 列的元素等于第一个因子的第 i 行诸元素与第二个因子的第 j 列诸相应元素的乘积之和．例如，

$$\begin{pmatrix} 2 & 1 & 3 \\ 3 & 1 & 1 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 2\cdot3+1\cdot1+3\cdot2 & 2\cdot1+1\cdot2+3\cdot4 \\ 3\cdot3+1\cdot1+1\cdot2 & 3\cdot1+1\cdot2+1\cdot4 \end{pmatrix}$$

$$= \begin{pmatrix} 13 & 16 \\ 12 & 9 \end{pmatrix}.$$

矩阵可以说是一个“复合”的对象，在它里面有很多元素出现，但是用一个字母来表示它而且用通常的符号来表示它的加法和乘法运算是有好处的，也是方便的．我们将用大写拉丁字母来表示矩阵．采用这种简写给矩阵论带来了简单性与概括性，即出现在

式子中的矩阵的元素所成的集合能在一些简短的与普通代数式子相似的式子里包含了某一数集所含数间的复杂关系式．例如，线性型组

$$a_{11}x_1+\cdots+a_{1n}x_n,$$
$$\cdots\cdots\cdots\cdots$$
$$a_{m1}x_1+\cdots+a_{mn}x_n$$

用矩阵记号可写作 AX，其中 A 是系数矩阵而 X 是由变数 $x_1, \cdots, x_n$ 组成的一"列"．线性方程组

$$a_{11}x_1+\cdots+a_{1n}x_n=b_1,$$
$$\cdots\cdots\cdots\cdots$$
$$a_{m1}x_1+\cdots+a_{mn}x_n=b_m$$

可以写作
$$AX=B,$$
其中 A 是系数矩阵，X 是变数列，B 是由自由项所组成的列．

基本运算——加法运算和乘法运算——并不是对于一切矩阵都可以定义．加法运算只对于同样类型的矩阵，即有同样行数和列数的矩阵，才有意义．而相加的结果也得到同样类型的一个矩阵．乘法运算只在第一个矩阵的列数等于第二个矩阵的行数时才有意义．相乘得到的矩阵的行数等于第一个因子的行数，而列数等于第二个因子的列数．

方阵的运算大部分服从数的运算规律，可是有些规律被破坏了．

我们列举矩阵运算的基本性质如下：

1．$A+B=B+A$（加法交换律）．

2．$(A+B)+C=A+(B+C)$（加法结合律）．

3．$c(A+B)=cA+cB$

3′．$(c_1+c_2)A=c_1A+c_2A$

（以上3、3′：以数乘矩阵的分配律．这里，c, c_1, c_2 是数而不是矩阵．）

4．$(c_1c_2)A=c_1(c_2A)$（以数乘矩阵的结合律）．

5．存在着"零"矩阵 $O=\begin{pmatrix}0 & \cdots & 0\\ \cdots & \cdots & \cdots\\ 0 & \cdots & 0\end{pmatrix}$，使得 $A+O=A$ 对任意

矩阵 A 成立.

6. $c\cdot O=O\cdot A=O$; 反之, 如果 $cA=O$, 则 $c=0$ 或 $A=O$(这里 c 是数).

7. 对于任意矩阵 A, 存在着一个反矩阵 $-A$, 使得 $A+(-A)=0$.

8. $(A+B)C=AC+BC$

8′. $C(A+B)=CA+CB$ (矩阵加法与乘法的分配律).

9. $(AB)C=A(BC)$ (乘法结合律).

10. $(cA)B=A(cB)=c(AB)$.

这些性质不仅对于方阵成立, 而且在上述公式中的每一个运算都有定义这个唯一条件下, 对于长方矩阵也成立. 对于同一阶的方阵, 这个条件自然满足.

上面所举的运算性质与数的运算性质相类似.

现在我们指出矩阵运算的两个特性.

第一, 对于矩阵的乘积, 即使是方阵的乘积, 交换律可能不成立, 即 AB 并不永远等于 BA. 例如:

$$\begin{pmatrix}3 & -2\\ -1 & 4\end{pmatrix}\begin{pmatrix}1 & 2\\ 3 & 2\end{pmatrix}=\begin{pmatrix}-3 & 2\\ 11 & 6\end{pmatrix},$$

$$\begin{pmatrix}1 & 2\\ 3 & 2\end{pmatrix}\begin{pmatrix}3 & -2\\ -1 & 4\end{pmatrix}=\begin{pmatrix}1 & 6\\ 7 & 2\end{pmatrix}.$$

第二, 我们都知道, 两个数的乘积等于零当且仅当这两个数中至少有一个等于零. 我们也知道, 这个定理在代数方程论里是很基本的. 可是两个矩阵的乘积可以等于零矩阵, 而这两个矩阵中没有一个等于零矩阵. 例如:

$$\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}1 & 1\\ -1 & -1\end{pmatrix}=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}.$$

我们再指出矩阵乘法的一个性质. 矩阵 A' 叫做矩阵 A 的**转置矩阵**, 如果 A' 中每一行的元素都顺序地换为矩阵 A 里相应的列的元素. 例如, 矩阵

$$A=\begin{pmatrix}1 & 2\\ 3 & 4\\ 5 & 6\end{pmatrix}$$

的转置矩阵是

$$A'=\begin{pmatrix}1 & 3 & 5\\ 2 & 4 & 6\end{pmatrix}.$$

矩阵的乘法运算与转置运算之间有关系式

$$(AB)'=B'A'.$$

根据矩阵乘法的规则很容易验证这个公式.

矩阵论是线性代数中重要的而且是不可缺少的部分，它在提出与解决线性代数的问题中起着工具的作用.

线性代数中的几何类比 除了上面提到的线性代数的概念和问题产生的源泉——数学分析的要求——之外，几何学，特别是解析几何学，也需要发展线性代数，而且用一些重要的概念和类比把线性代数充实起来. 大家都知道，平面解析几何以及在很大程度上空间解析几何在关于直线和平面的理论里，在最简单的形式下，利用着线性代数这个工具. 实际上，平面上的直线由一个含两个变数的线性方程所给出. 空间中的平面由一个含三个变数的线性方程给出，而空间中的直线由两个线性方程给出.

大家都知道，如果采用向量的概念，可以使解析几何特别简单和清楚，因此也使得最简单的线性方程组的理论特别的简单和清楚. 采用某种广义的向量的概念也使得线性代数同样的简单和清楚. 推广的途径如下(空间中的)：向量由三个数给出，这三个数是它在坐标轴上的射影. 三个实数所组成的数组也可以在几何上用(空间中的)向量表示.

可以定义两个向量的加法运算(“按平面四边形规律”)和乘以数的运算. 这些运算按照被表成向量的力、速度、加速度和其他的物理量的类似运算来定义.

如果向量由它的坐标(即它在坐标轴上的射影)给出，那么在向量上进行的加法运算和乘以数的运算就相应着由它的坐标所组

成的行(或列)上同名的运算.

这样一来，由三个元素组成的行或列就宜于几何地解释作三维空间中的向量，同时在“行”(或“列”)上所进行的运算就解释作在空间中向量上所进行的相应的运算，使得由三个元素所组成的行(或列)的代数在形式上与三维空间中的向量代数没有差别. 这种情况使我们很自然地在线性代数中引进几何术语.

n 个数$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$组成的列(或行)可以看作是“向量”，即“$n$ 维向量空间”的一个元素. 向量$\begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{pmatrix}$认为是向量$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$和$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$的和; 向量$\begin{pmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{pmatrix}$认为是向量$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$乘以数 c 的乘积. 按定义, 所有向量(列)的全体组成 n 维算术向量空间.

与此同时, 还可以引进 n 维点空间的概念, 它使 n 个实数组成的列对应于一个几何图像——点. 这时 n 维向量空间可以如下定义: 对于每一对点 A 和 B, 对应着一个从点 A 出发到点 B 为止的向量, 把点 B 和点 A 的相应坐标之差按定义规定是这个向量的坐标(在坐标轴上的射影). 两个向量看作是相等的, 如果它们的坐标相等. 这正像三维空间中我们把两个向量看作是相等的, 如果其中之一可以从另一个经平行移动而得到.

很自然地可以建立 n 维向量空间与 n 维点空间之间的一个一一对应关系.

点$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$被取作“坐标原点”，每一个点都让一个向量和它相

应，这就是从坐标原点出发到这个点为止的向量。这时每一个向量也都与一个点，即这个向量的端点相应，其中假设这个向量的出发点是坐标原点。点空间的引进所建立的新的类比，使我们能更好地"观察"n维空间。

但是在更进一步推广时(§ 2)，点空间的严格定义变得相当复杂，因此我们不准备采用这个概念。愿意采用从考虑点空间而产生的类比的读者，需要把向量空间中的元素看作从坐标原点出发的向量。

几何术语的引进使我们有可能在线性代数中利用基于几何直观的类比，而这些几何直观是在研究三维空间的几何时形成的。

当然需要很小心地采用这种类比，要估计到能够严格地验证几何直观论证的可能性并只采用几何概念的确切定义以及严格证明了的定理。

n维向量空间中元素的特性是加法和乘以数的乘法运算的存在性，而这些运算的性质就像数所进行的运算那样，即正像在描述矩阵运算的性质时所指出的。对于加法运算，交换律和结合律成立，分配律(当乘以数时)也成立；加法运算是单值可逆的，向量乘以数的乘积当且仅当这个向量是零向量或这个数等于零时等于零向量。

不仅列(及行)具备上面所说的这些特性，就是同一类型的矩阵集合以及物理向量：力、速度、加速度等等也具备这些性质。完全是另外性质的数学对象，如一个变数的多项式的全体、已知区间$[a, b]$上的连续函数的全体、线性齐次微分方程的解的全体等等，也都具备这些性质。

上面举出的例子指出，进一步推广向量空间的概念，即引进一般的线性空间，是有益的。这种广义空间中的元素可以是任意数学对象或物理对象，但对于它们，可以用某种自然方式来定义加法和乘以数的乘法。我们将会看到，过渡到线性空间概念的这样一个一般而抽象的过程并不带来任何理论上的复杂性：任何线性空间(当然是指n维的，这个概念将在下一节里解释清楚)在结构上

和性质上与算术线性空间没有什么两样，但是这样推广之后应用的范围却扩大了，运用线性代数的方法到很广阔的自然科学理论问题上的可能性也出现了．

§2. 线 性 空 间

线性空间的定义 我们现在来给出线性空间的严格定义．

一组任一种类对象的集合称为一个线性空间．如果在其中和及乘以数的乘积有意义而且满足下列诸要求：

1．$(\boldsymbol{X}+\boldsymbol{Y})+\boldsymbol{Z}=\boldsymbol{X}+(\boldsymbol{Y}+\boldsymbol{Z})$．

2．存在着“零元素”$\mathbf{0}$，使得对任意 $\boldsymbol{X}$，$\boldsymbol{X}+\mathbf{0}=\boldsymbol{X}$．

3．对于任意元素 $\boldsymbol{X}$，存在着一个逆元素 $-\boldsymbol{X}$，使得 $\boldsymbol{X}+(-\boldsymbol{X})=\mathbf{0}$．

4．$\boldsymbol{X}+\boldsymbol{Y}=\boldsymbol{Y}+\boldsymbol{X}$．

5．$1\cdot\boldsymbol{X}=\boldsymbol{X}$．

6．$c_1(c_2\boldsymbol{X})=c_1c_2\boldsymbol{X}$．

7．$(c_1+c_2)\boldsymbol{X}=c_1\boldsymbol{X}+c_2\boldsymbol{X}$．

8．$c(\boldsymbol{X}+\boldsymbol{Y})=c\boldsymbol{X}+c\boldsymbol{Y}$．

这里 $\boldsymbol{X}$，$\boldsymbol{Y}$，$\boldsymbol{Z}$ 是线性空间里的元素而 1，c_1，c_2，c 是数．

这些要求（也称为线性空间的公理）是很自然的而且是加法及乘以数的乘法运算的基本性质的形式叙述．这些性质与在任意广义意义下所理解的这些运算的概念不可分地联系在一起．有着某种物理意义的运算说成是加法和乘以数的乘法，恰是当这些运算适合条件 1—8 时．

下面指出这些公理的某些推论：

1．空间的零元素 $\mathbf{0}$ 是唯一的，即只存在一个适合公理 2 的元素．

2．对于已知元素 $\boldsymbol{X}$，存在着唯一的一个逆元素．

3．“减法”是有意义的，即如果和及一个加项已知，则可唯一确定第二个加项；即，如果 $\boldsymbol{X}+\boldsymbol{Z}=\boldsymbol{Y}$，则 $\boldsymbol{Z}=\boldsymbol{Y}+(-\boldsymbol{X})$．

4. $\mathbf{0}\cdot X=c\cdot\mathbf{0}=\mathbf{0}$.

5. 如果 $cX=\mathbf{0}$, 则 $c=0$ 或 $X=\mathbf{0}$.

6. $-X=(-1)X$.

这些推论的证明是很简单的，我们不准备证明它们. 今后我们把线性空间的元素称为向量.

向量的线性相关与线性无关 现在我们转到向量的线性相关与线性无关这个重要概念上来.

向量 $c_1X_1+c_2X_2+\cdots+c_mX_m$ 称为向量 $X_1, X_2, \cdots, X_m$ 以数值 $c_1, c_2, \cdots, c_m$ 为系数的线性组合. 如果在向量 $X_1, X_2, \cdots, X_m$ 中能找出一个向量是其余向量的线性组合，则向量 $X_1, X_2, \cdots, X_m$ 称为线性相关. 如果向量 $X_1, X_2, \cdots X_m$ 中没有一个是其余向量的线性组合, 则向量 $X_1, X_2, \cdots, X_m$ 称为线性无关.

不难看出, 向量 $X_1, X_2, \cdots, X_m$ 线性无关的充分必要条件是关系式 $c_1X_1+c_2X_2+\cdots+c_mX_m=\mathbf{0}$, 仅当 $c_1=c_2=\cdots=c_m=0$ 时.

对于通常三维空间里的向量，线性相关和线性无关的概念有着简单的几何意义.

设有两个向量 X_1 和 X_2. 它们线性相关是说其中的一个向量是另一个向量的"线性组合", 即只与这个向量差一个常数因子. 这就是说, 这两个向量落在同一条直线上, 即它们的方向相同或相反.

反之，如果两个向量落在同一条直线上，那么它们线性相关. 因之, 两个向量 X_1 和 X_2 线性无关是说这两个向量不能落在一条直线上; 它们的方向根本不同.

现在来考察三个向量的线性相关与线性无关. 设向量 X_1, X_2 和 X_3 线性相关, 为了确定起见, 假设向量 X_3 是向量 X_1 和 X_2 的线性组合. 显然这时 X_3 落在含向量 X_1 和 X_2 的平面里，即 X_1, X_2, X_3 这三个向量落在同一张平面里. 容易看到，如果向量 X_1, X_2, X_3 落在同一张平面里, 它们就线性相关. 实际上, 如果向量 X_1 和 X_2 不落在同一条直线上, 那么 X_3 可以按 X_1 和 X_2 来分解, 即将 X_3 表作 X_1 和 X_2 的线性组合的形状. 如果 X_1 和 X_2 落

在同一条直线上，那么 X_1 和 X_2 已经线性相关.

这样一来，三个向量 X_1, X_2, X_3 线性相关与它们落在同一张平面上等价. 因此，X_1, X_2, X_3 线性无关当且仅当它们不落在同一张平面上.

三维空间中的四个向量永远线性相关. 实际上，如果向量 X_1, X_2, X_3 线性相关，那么对于任意 X_4，向量 X_1, X_2, X_3, X_4 都线性相关. 如果 X_1, X_2, X_3 线性无关，那么它们不落在同一张平面上而且任何向量 X_4 都可以按 X_1, X_2, X_3 分解，即表成它们线性组合的形状.

上述讨论可以作如下的概括：

三维空间中的向量 X_1, X_2, …, X_k($k\geqslant 3$)线性相关当且仅当它们落在维数小于 k 的空间(直线或平面)之中.

将来严格地定义了子空间和维数之后，在一般情形，向量 X_1, X_2, …, X_k 线性相关等价于它们落在维数小于 k 的一个子空间中，即线性相关的"几何意义"与三维空间的情形仍然相同.

在线性空间的理论中，下面这个定理起着重要的作用. 如果向量 X_1, X_2, …, X_m 中每一个都是向量 Y_1, Y_2, …, Y_k 的线性组合，而且 $m>k$，则向量 X_1, X_2, …, X_m 线性相关(关于线性组合线性相关性的定理).

当 $k=1$ 时，定理显然成立. 对于 $k>1$，定理可对 k 用数学归纳法来证明.

空间的基与维数 在三维空间中，任意三个不落在同一张平面里的向量 X_1, X_2, X_3(即线性无关的三个向量)组成这个空间的一组基，这就是说，任意向量皆可按向量 X_1, X_2, X_3 来分解，即可表作它们的线性组合.

一般的线性向量空间显然可以分成两类.

可能在空间中存在着任意大数目的线性无关向量. 这种空间称为无限维空间，对于它的研究超出了线性代数的范围，而是特殊的数学分支——泛函分析(见第十九章)——的对象.

线性空间称为有限维的，如果在它里面存在着线性无关向量

的个数的一个有限上界，即这样的一个数 n，使得在空间中存在着 n 个线性无关的向量，而任意个数大于 n 的向量组都线性相关. n 这个数称为空间的**维数**.

例如，通常三维几何空间里的向量所组成的空间按照上面的定义来说是三维的. 实际上，在三维几何空间中存在着三个线性无关的向量组，而任意四个向量都线性相关.

n 数列所组成的空间在上面的定义下是 n 维的. 实际上，在这个空间里存在着 n 个线性无关的向量，例如，

$$e_1=\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix},\ e_2=\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix},\ \cdots,\ e_n=\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}, \tag{2}$$

而且这个空间里任何一个向量 $\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$ 都是它们的线性组合，即 $x_1e_1+x_2e_2+\cdots+x_ne_n$. 因此，根据关于线性组合线性相关性的定理，任意一组个数大于 n 的向量都线性相关.

一个变数的多项式的全体组成一个线性空间. 实际上，对于多项式可以按照通常的方法来定义加法和乘以数的乘法，它们适合公理 1—8. 然而，这个空间是无限维的，因为对于任意 N，向量 $1, x, \cdots, x^N$ 都线性无关. 次数不超过已知数 N 的多项式的集合组成一个有限维向量空间，其维数是 $N+1$. 每一个次数不超过 N 的多项式都是 $1, x, \cdots, x^N$ 的线性组合，而根据关于线性相关性的那个定理，任意一组个数大于 $N+1$ 的次数 $\leqslant N$ 的多项式都线性相关.

现在我们对于 n 维空间来引进**基**这个重要概念. 基就是空间中这样一组线性无关向量的集合，使得空间中任意一个向量都是这个集合中向量的线性组合. 例如，在数列空间中，向量集合 (2) 就是一组基. 在次数 $\leqslant N$ 的多项式组成的空间中，可以取“向量”

$1, x, \cdots, x^N$ 作为一组基. 在三维几何空间中，任意三个线性无关的向量组起着基的作用.

在 n 维线性空间中，任意 n 个线性无关向量的集合（根据 n 维空间的定义，这样一组向量是存在的）组成这个空间的一组基. 实际上，设 $e_1, e_2, \cdots, e_n$ 是 n 维线性空间中线性无关的向量而 $\boldsymbol{X}$ 是这个空间里的任意一个向量，那么向量 $\boldsymbol{X}, e_1, \cdots, e_n$ 线性相关（因为它们的个数大于 n），即可以找到一组不全等于 0 的数 $c, c_1, c_2, \cdots, c_n$ 使得 $c\boldsymbol{X}+c_1e_1+\cdots+c_ne_n=0$. 这时，$c\neq 0$，因为如果 $c=0$，那么向量 $e_1, e_2, \cdots, e_n$ 就会线性相关. 因此 $\boldsymbol{X}=-\frac{c_1}{c}e_1-\cdots-\frac{c_n}{c}e_n$，即空间中的任意向量都是 $e_1, e_2, \cdots, e_n$ 这 n 个向量的线性组合.

n 维线性空间中任意一组基全都由 n 个向量组成. 实际上，基里的向量是线性无关的，因而它的个数不能大于 n. 另一方面，设 $e_1, e_2, \cdots, e_k$ 是 n 维空间中的任意一组基，我们已经证明了 $k\leqslant n$. 又根据基的定义，空间中任何向量都是向量 $e_1, e_2, \cdots, e_k$ 的线性组合；再根据关于线性相关性的定理，任意一组个数大于 k 的向量都线性相关，因此空间的维数 n 不能大于基里向量的个数 k. 于是 $k=n$，这就是所要证明的.

现在来引进向量对于已知的一组基 $e_1, e_2, \cdots, e_n$ 的**坐标**. 上面已经说过，任意向量 $\boldsymbol{X}$ 都是基里向量的线性组合. 这种表示法是唯一的. 实际上，设向量 $\boldsymbol{X}$ 以两种方式用基 $e_1, e_2, \cdots, e_n$ 表出：

$$\boldsymbol{X}=x_1e_1+x_2e_2+\cdots+x_ne_n,$$
$$\boldsymbol{X}=x'_1e_1+x'_2e_2+\cdots+x'_ne_n.$$

那么 $(x_1-x'_1)e_1+(x_2-x'_2)e_2+\cdots+(x_n-x'_n)e_n=0$，因此，根据 $e_1, e_2, \cdots, e_n$ 的线性无关性就推出 $x_1=x'_1, \cdots, x_n=x'_n$.

任意向量 $\boldsymbol{X}$ 表成基中向量的线性组合时的系数 $x_1, x_2, \cdots, x_n$ 称为向量 $\boldsymbol{X}$ 对于这组基的**坐标**. 于是，**只要选定了空间的基之后**，任意向量都与它的坐标所组成的行（或列）相应，反之，任意 n 个数所组成的行（或列）都可以看作某个向量的坐标.

向量的加法运算与乘以数的运算也分别对应着坐标组成的行(或列)的同名的运算.

因此,任意 n 维线性空间,如不论它元素的性质(它们可能是函数,可能是矩阵,也可能是某种物理量等等),对于这些运算说来与行(或列)空间并没有什么差别. 因此,正像上面所说,一般的、公理化的线性空间的概念与行空间的概念相比并没有带来什么复杂性,可是却大大地扩展了这个概念的应用范围.

两个元素组对于给定的一组运算(或元素间的其他关系)来说,性质的一致性在数学上称为同构. 代数系统同构的确切定义将在第二十章给出. 利用这个名词我们可以说,一切 n 维线性空间,不论其元素的性质如何,是彼此同构的,而且同构于唯一的一个模型——行空间.

子空间 n 维线性空间 R_n 中的一个向量集合称为这个空间的一个子空间,如果这个集合中的任意个向量的任何线性组合都仍然属于这个集合之中. 显然,空间 R_n 的子空间本身也是线性空间,因而有基与维数. 同样也显然,子空间的维数不能超过整个空间的维数,而它们当且仅当子空间与整个空间相重时可能相等.

三维向量空间中落在一张平面或一条直线上的向量的集合就是子空间的例子.

最常见的是由一组向量"张成"的子空间. 这种子空间定义如下:设给了空间 $\boldsymbol{R}_n$ 中一组线性无关或线性相关的向量 $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, …, $\boldsymbol{X}_m$. 那么这些向量的线性组合的全体 $\{c_1\boldsymbol{X}_1+c_2\boldsymbol{X}_2+\cdots+c_m\boldsymbol{X}_m\}$ 组成 $\boldsymbol{R}_n$ 的一个子空间,它称为由向量 $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, …, $\boldsymbol{X}_m$ 张成的子空间.

这种子空间的维数称为向量组 $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, …, $\boldsymbol{X}_m$ 的秩. 容易看出,一组向量的秩等于这组向量里所含的线性无关向量的最大数.

仅由零向量所组成的"集合",形式地满足子空间所要求的条件. 这个子空间的维数被认为是零.

如果给了空间 $\boldsymbol{R}_n$ 的两个子空间,那么从它们自然还可以造

出两个子空间——它们的向量和与它们的交.

两个子空间 P 和 Q 的向量和是属于空间 P 的向量与属于空间 Q 的向量的一切和的集合. 向量和可以看作是子空间 P 的基和 Q 的基的并集所张成的子空间.

两个子空间的交是同时属于这两个子空间的一切向量的集合. 例如,通常三维空间中,两个平面(即两个二维向量子空间)的向量和是整个空间(如果这两个平面不相重), 而它们的交是一条直线(也当它们不相重时).

两个已知子空间的维数 p 和 q, 与它们的向量和的维数 t 以及交的维数 s, 适合下面这个有趣的关系式:

$$p+q=t+s.$$

这个断言的证明从略.

从这个关系式我们可以作出一些在特殊情况下子空间的交的推论来. 例如, 在四维空间中两张不相重的平面(即二维子空间)一般说来只交于一点(它们的交的维数等于零), 而两张平面只在它们的向量和是三维时, 即如果这两张平面落在某个三维子空间中时,交于一条直线. 实际上,这时 $t+s=2+2=4$, 由此推出仅当 $t=3$ 时 $s=1$.

复线性空间 在叙述行空间及一般线性空间的时候, 我们并没有确定在定义向量乘以数的运算时谈的是什么数. 因为我们是从推广通常的向量,即三维几何空间中的方向线段出发的,所以我们心目中的数是任意实数. 这样构成的线性空间称为实线性空间, 是通常三维向量空间的最自然的推广. 然而复线性空间的研究对于近代数学中的很多问题是有用的. 所谓复线性空间是指一组元素的集合,在其中定义了加法和乘以复数的乘法,而且这些运算适合所有公理 1—8. 行的元素是复数的行空间是复线性空间的例子.

复空间的理论在形式上与实空间的理论并没有什么基本不同.

然而即使是二维复空间也没有直观的几何解释. 问题在于 n

维复空间也可以看作是实空间，因为既然对于它的元素定义了乘以复数的乘法，自然也就定义了乘以实数的乘法．但将 n 维复空间看作实空间时，它的维数是 $2n$，即大了一倍．实际上，如果 e_1, e_2, …, e_n 是复空间的一组基，那么将它看作实空间时可以取向量

$$e_1,\ ie_1,\ e_2,\ ie_2,\ \cdots,\ e_n,\ ie_n,\ \text{这里 } i=\sqrt{-1},$$

作为这个空间的一组基．

因此，二维复空间可以解释作实的、但是是四维的空间．

更进一步，如果选取任意一个数集，它异于一切实数的集合及一切复数的集合，把它作为空间中"向量"乘以数的乘法所容许的数的集合，只要在这个数集中进行基本算术运算——加、减、乘、除——的结果仍然属于这个集合，那么线性空间的理论形式上也没有什么改变．适合上面所说的条件的数集称为数域(这个概念将会在第二十章中更详细地予以研究)．譬如，有理数集就是数域的一个例子．

在一些接近数论的代数分支中，有效地运用了任意域上线性空间的理论．

n 维欧氏空间 在前面的讨论里，我们还没有推广通常向量空间中的一些重要的概念，特别是向量长度和向量夹角的概念．我们知道在解析几何中，在关于直线与圆的交点，平行性以及很多别的问题里，并没有用到这些概念．空间的性质，在描述它们的时候并不需要向量长度和夹角这些概念的，可以刻划为在任意仿射变换(见第一卷第三章 § 11)之下不变的性质．根据这个原则，线性空间中没有定义向量长度这个概念的称为仿射空间．

然而在很多数学问题里需要将向量长度及夹角的概念推广到 n 维空间中来．这可以用类似于平面上或空间中向量理论的方法来进行推广．

先来研究实的行向量空间．向量 $\boldsymbol{X}=(x_1,\ x_2,\ \cdots,\ x_n)$ 的长度定义为 $|\boldsymbol{X}|=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}$ 这个数．这是很自然的，因为当 $n=2$ 和 $n=3$ 时，正是按照这个公式从向量对于笛卡儿坐标轴的坐标来计算它的长度的．

向量的夹角这个概念，也可以很自然地从下面的考察引进来。在平面上或空间中，向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的夹角就是以 $AB=|\boldsymbol{X}|$，$AC=|\boldsymbol{Y}|$ 和 $BC=|\boldsymbol{X}-\boldsymbol{Y}|$ 为边的三角形中 A 的顶角。

在 n 维空间中自然地采用它作为向量的夹角，即可以这样想象：从 n 维空间中把这一对向量"抽出"来再把它们"安置"在平面中而不改变它们的长度和夹角。然而在这样一个定义中却有不严格之处，即以向量长度 $|\boldsymbol{X}|$，$|\boldsymbol{Y}|$ 和 $|\boldsymbol{X}-\boldsymbol{Y}|$ 为边的三角形的存在性还需要证明。

为了免除这个不确切之处，我们引进计算角的公式。从三角中的著名公式

$$BC^2=AB^2+AC^2-2AB\cdot AC\cos\varphi$$

可以推出

$$\cos\varphi=\frac{|\boldsymbol{X}|^2+|\boldsymbol{Y}|^2-|\boldsymbol{X}-\boldsymbol{Y}|^2}{2|\boldsymbol{X}|\cdot|\boldsymbol{Y}|}$$

$$=\frac{x_1^2+\cdots+x_n^2+y_1^2+\cdots+y_n^2+(x_1-y_1)^2+\cdots+(x_n-y_n)^2}{2|\boldsymbol{X}|\cdot|\boldsymbol{Y}|}$$

$$=\frac{x_1y_1+\cdots+x_ny_n}{|\boldsymbol{X}|\cdot|\boldsymbol{Y}|}.$$

如果像在三维空间中一样，我们保留两个向量的"数量积"这个名词为它们夹角的余弦乘以它们长度的乘积，那么向量的数量积可以按照公式

$$\boldsymbol{X}\cdot\boldsymbol{Y}=x_1y_1+\cdots+x_ny_n$$

来计算，这个公式当 $n=2$ 和 $n=3$ 时与通常向量的数量积的著名公式一致。

严格地说，必须把式子 $x_1y_1+\cdots+x_ny_n$ 取作向量积的定义（因为利用角来定义向量积是有不严格之处的），然后再用公式

$$\cos\varphi=\frac{\boldsymbol{X}\cdot\boldsymbol{Y}}{|\boldsymbol{X}|\cdot|\boldsymbol{Y}|}\tag{3}$$

来定义向量的夹角。

我们就这样来做。

为了证明角的这个定义的合理性，我们必须证明(3)式右方的

绝对值不超过 1，即 $(\boldsymbol{X}\cdot\boldsymbol{Y})^2\leqslant|\boldsymbol{X}|^2\cdot|\boldsymbol{Y}|^2$.

将此不等式展开，即得

$$(x_1y_1+\cdots+x_ny_n)^2\leqslant(x_1^2+\cdots+x_n^2)(y_1^2+\cdots+y_n^2).$$

这个不等式称为柯西-布涅考夫斯基不等式，而且是可以直接证明的，只是计算相当复杂. 我们用下面这个间接的方法来证明.

首先我们指出数量积有下面这些性质：

1'. 当 $\boldsymbol{X}\neq0$ 时，$\boldsymbol{X}\cdot\boldsymbol{X}=|\boldsymbol{X}|^2>0$.

2'. $\boldsymbol{X}\cdot\boldsymbol{Y}=\boldsymbol{Y}\cdot\boldsymbol{X}$.

3'. $(c\boldsymbol{X})\cdot\boldsymbol{Y}=c(\boldsymbol{X}\cdot\boldsymbol{Y})$.

4'. $(\boldsymbol{X}_1+\boldsymbol{X}_2)\cdot\boldsymbol{Y}=\boldsymbol{X}_1\cdot\boldsymbol{Y}+\boldsymbol{X}_2\cdot\boldsymbol{Y}$.

这些性质的成立可以从用坐标表出数量积的式子推出.

现在引进向量 $\boldsymbol{Y}+t\boldsymbol{X}$ 来研究它，其中 t 是任意实数. 我们是 $|\boldsymbol{Y}+t\boldsymbol{X}|^2\geqslant0$，因为向量长度的平方不能是负的. 但根据数量积的定义，$|\boldsymbol{Y}+t\boldsymbol{X}|^2=|\boldsymbol{Y}|^2+2t\boldsymbol{X}\cdot\boldsymbol{Y}+t^2|\boldsymbol{X}|^2$. 再者我们知道，二次三项式对于实变数 t 的一切值当且仅当它的根是虚数或相等，即它的判别式是负数或等于零时永远取非负值. 但二次三项式 $|\boldsymbol{Y}|^2+2t\boldsymbol{X}\cdot\boldsymbol{Y}+t^2|\boldsymbol{X}|^2$ 的判别式等于 $4(\boldsymbol{X}\cdot\boldsymbol{Y})^2-4|\boldsymbol{X}|^2|\boldsymbol{Y}|^2$，因此 $(\boldsymbol{X}\cdot\boldsymbol{Y})^2-|\boldsymbol{X}|^2|\boldsymbol{Y}|^2\leqslant0$，此式等价于柯西-布涅考夫斯基不等式.

从所证明的不等式推出 $\dfrac{|\boldsymbol{X}\cdot\boldsymbol{Y}|}{|\boldsymbol{X}||\boldsymbol{Y}|}\leqslant1$，因此按照 (3) 式来定义夹角是合理的.

更进一步，不难导出不等式

$$|\boldsymbol{X}|-|\boldsymbol{Y}|\leqslant|\boldsymbol{X}\pm\boldsymbol{Y}|\leqslant|\boldsymbol{X}|+|\boldsymbol{Y}|,$$

由此更不难推出以 $|\boldsymbol{X}|$，$|\boldsymbol{Y}|$ 和 $|\boldsymbol{X}-\boldsymbol{Y}|$ 为边的三角形的存在性. 因此，前面所给出不严格的、但几何上却很直观的角的定义也获得了合理的基础.

n 维欧氏空间的公理定义 在前面，我们在行空间中引进了向量的长度、夹角及数量积的概念. 在一般的用公理定义的 n 维实线性空间中，这些概念也可以用公理来定义，而基础是数量积这

个概念.

线性实空间中向量的数量乘法是对于任意一对向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 相应一个实数的对应，这个相应的实数称为它们的数量积 $\boldsymbol{X}\cdot\boldsymbol{Y}$，而且这个对应应当适合下面这些要求（公理）：

1′. 当 $\boldsymbol{X}\neq\mathbf{0}$ 时，$\boldsymbol{X}\cdot\boldsymbol{X}>0$；$\mathbf{0}\cdot\mathbf{0}=0$.

2′. $\boldsymbol{X}\cdot\boldsymbol{Y}=\boldsymbol{Y}\cdot\boldsymbol{X}$.

3′. $(c\boldsymbol{X})\cdot\boldsymbol{Y}=c(\boldsymbol{X}\cdot\boldsymbol{Y})$.

4′. $(\boldsymbol{X}_1+\boldsymbol{X}_2)\cdot\boldsymbol{Y}=\boldsymbol{X}_1\cdot\boldsymbol{Y}+\boldsymbol{X}_2\cdot\boldsymbol{Y}$.

然后把数 $\sqrt{\boldsymbol{X}\cdot\boldsymbol{X}}$ 取作向量 $\boldsymbol{X}$ 的长度，而数 $\dfrac{\boldsymbol{X}\cdot\boldsymbol{Y}}{|\boldsymbol{X}|\cdot|\boldsymbol{Y}|}$ 就取作向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的夹角. 为了证明后面这个定义的合法性，需要证明柯西-布涅考夫斯基不等式 $(\boldsymbol{X}\cdot\boldsymbol{Y})^2\leqslant|\boldsymbol{X}|^2|\boldsymbol{Y}|^2$ 成立. 但这可以完全像前面一样来证明. 在刚才的证明里，只用到数量积的性质 1′, 2′, 3′, 4′，而行空间的特殊性在证明中并不起作用. 在实线性空间中引进了适合公理 1′—4′ 的数量积的就称为欧氏空间.

在数学中研究了很多具体的线性空间，它们的数量积可以用种种不同的方法引进，它们的选取决定于问题的实质. 例如，$\boldsymbol{X}(t)$ 是定义在已知区间 $a\leqslant t\leqslant b$ 上的单变数函数. 在以 $\boldsymbol{X}(t)$ 为元素的空间中，两个元素 $\boldsymbol{X}(t)$ 和 $\boldsymbol{Y}(t)$ 的数量积常取作数 $\int_a^b\boldsymbol{X}(t)\boldsymbol{Y}(t)dt$ 或数 $\int_a^b\boldsymbol{X}(t)\boldsymbol{Y}(t)p(t)dt$，其中 $p(t)$ 是某个正函数. 容易看出，无论那一个定义都满足公理 1′—4′.

正交性 标准正交基 欧氏空间中两个向量称为正交（或垂直），如果它们的数量积等于零. 容易看出，两两正交的一组非零向量永远线性无关. 实际上，设 $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_m$ 是一组两两正交的非零向量，并设 $c_1\boldsymbol{X}_1+c_2\boldsymbol{X}_2+\cdots+c_m\boldsymbol{X}_m=0$，那么根据数量积的性质就有 $\boldsymbol{X}_1(c_1\boldsymbol{X}_1+c_2\boldsymbol{X}_2+\cdots+c_m\boldsymbol{X}_m)=c_1|\boldsymbol{X}_1|^2=0$. 由此推出 $c_1=0$. 用同样的方法可以证明 $c_2=\cdots=c_m=0$. 因此，$\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_m$ 线性无关.

从上面证明的事实推出，n 维空间中顶多存在 n 个两两正交的非零向量，而任一由 n 个两两正交非零向量所组成的集合都是空间的一组基。此外，如果 n 个两两正交的向量的长度都等于 1，那么它们所组成的基就称为标准正交基。

不难证明，在欧氏空间中存在着标准正交基，而它们的个数还可以无限多。但我们不准备证明这个事实。此外，如果在空间 R 中选出了某个子空间，那么这个子空间的标准正交基可以添加上一些向量补充成整个空间的一组标准正交基。

欧氏空间中的向量最好用它在某个标准正交基中的坐标给出，因为这时数量积的式子特别简单。实际上，如果向量 $\boldsymbol{X}$ 在标准正交基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 中的坐标是 $(x_1, x_2, \cdots, x_n)$，而向量 $\boldsymbol{Y}$ 的坐标是 $(y_1, y_2, \cdots, y_n)$，即

$$\boldsymbol{X}=x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+\cdots+x_n\boldsymbol{e}_n \text{ 及 } \boldsymbol{Y}=y_1\boldsymbol{e}_1+y_2\boldsymbol{e}_2+\cdots+y_n\boldsymbol{e}_n,$$

那么根据数量积的性质就有

$$\begin{aligned}\boldsymbol{X}\cdot\boldsymbol{Y}=&x_1y_1\boldsymbol{e}_1\cdot\boldsymbol{e}_1+x_1y_2\boldsymbol{e}_1\cdot\boldsymbol{e}_2+\cdots+x_1y_n\boldsymbol{e}_1\cdot\boldsymbol{e}_n\\&+x_2y_1\boldsymbol{e}_2\cdot\boldsymbol{e}_1+x_2y_2\boldsymbol{e}_2\cdot\boldsymbol{e}_2+\cdots+x_2y_n\boldsymbol{e}_2\cdot\boldsymbol{e}_n+\\&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\&+x_ny_1\boldsymbol{e}_n\cdot\boldsymbol{e}_1+x_ny_2\boldsymbol{e}_n\cdot\boldsymbol{e}_2+\cdots+x_ny_n\boldsymbol{e}_n\cdot\boldsymbol{e}_n\\=&x_1y_1+x_2y_2+\cdots+x_ny_n,\end{aligned}$$

因为当 $i\neq k$ 时，$\boldsymbol{e}_i\cdot\boldsymbol{e}_k=0$，而当 $i=1, 2, \cdots, n$ 时，$\boldsymbol{e}_i\cdot\boldsymbol{e}_i=|\boldsymbol{e}_i|^2=1$。特别有 $\boldsymbol{X}\cdot\boldsymbol{X}=x_1^2+x_2^2+\cdots+x_n^2$。

因此，向量的长度和数量积可以按行空间中同样的公式用它在标准正交基中的坐标表出。

从欧氏空间的一个模型——行空间——到一般的、用公理定义的欧氏空间的过程并没有引起任何复杂化，而只是扩大了理论的应用范围。

现在还要研究关于向量在子空间上的正交投影问题。设 R_n 是一个 n 维欧氏空间，P_m 是它的一个 m 维子空间；再设 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_m, \boldsymbol{f}_1, \cdots, \boldsymbol{f}_{n-m}$ 是 R_n 的一组标准正交基，并包有子空间 P_m 的一组标准正交基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_m$。由向量 $\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_{n-m}$ 所

张成的子空间 Q_{n-m} 称为 P_m 的**正交补空间**，它的维数等于 $n-m$. 正交补空间 Q_{n-m} 可以刻划为与子空间 P_m 中一切向量都正交的全体向量所组成的集合.

R_n 中的任意向量 $\boldsymbol{Z}$ 可以唯一地表作 P_m 中一个向量 $\boldsymbol{X}$ 和 Q_{n-m} 中一个向量 $\boldsymbol{Y}$ 的和. 从向量 $\boldsymbol{Z}$ 可以唯一表成

$$\boldsymbol{Z}=x_1\boldsymbol{e}_1+\cdots+x_m\boldsymbol{e}_m+y_1\boldsymbol{f}_1+\cdots+y_{n-m}\boldsymbol{f}_{n-m}$$

的形状就知道这是很显然的，因为令 $\boldsymbol{X}=x_1\boldsymbol{e}_1+\cdots+x_m\boldsymbol{e}_m$，$\boldsymbol{Y}=y_1\boldsymbol{f}_1+\cdots+y_{n-m}\boldsymbol{f}_{n-m}$ 就行了.

向量 $\boldsymbol{X}$ 称为向量 $\boldsymbol{Z}$ 在 P_m 上的**正交投影**.

酉空间* 向量的长度和数量积的概念，在复空间中也可以定义. 我们仍然以数量积的概念作为基础，它可以如下地定义：让复空间中每一对向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 都相应一个复数(不必要是实数)，称为它们的数量积 $\boldsymbol{X}\cdot\boldsymbol{Y}$. 数量乘法这个运算应该适合下列公理：

1″. 若 $\boldsymbol{X}\neq 0$，$\boldsymbol{X}\cdot\boldsymbol{X}$ 是正实数，而 $\boldsymbol{0}\cdot\boldsymbol{0}=0$.

2″. $\boldsymbol{Y}\cdot\boldsymbol{X}=(\boldsymbol{X}\cdot\boldsymbol{Y})'$，这里撇表示复共轭数.

3″. $(c\boldsymbol{X})\cdot\boldsymbol{Y}=c(\boldsymbol{X}\cdot\boldsymbol{Y})$ 对任意复数 c.

4″. $(\boldsymbol{X}_1+\boldsymbol{X}_2)\cdot\boldsymbol{Y}=\boldsymbol{X}_1\cdot\boldsymbol{Y}+\boldsymbol{X}_2\cdot\boldsymbol{Y}$ (分配律).

在以复数为元素的行所组成的行空间中，向量 $(x_1, x_2, \cdots, x_n)$ 和 $(y_1, y_2, \cdots, y_n)$ 的数量积可以取为 $x_1y_1'+x_2y_2'+\cdots+x_ny_n'$ 这个数. 容易证明，这时公理 1″—4″ 都成立.

取 $\sqrt{\boldsymbol{X}\cdot\boldsymbol{X}}$ 这个数作为向量的长度. 向量间夹角的概念不去定义.

复线性空间如具有适合公理 1″—4″ 的数量积，则称为**酉空间**.

§3. 线性方程组

两个含两个未知数的方程组及三个含三个未知数的方程组

两个含两个未知数的方程组的一般形状是

* 也可译作单式正交空间. ——译者注

$$a_1x+b_1y=c_1,$$
$$a_2x+b_2y=c_2.$$

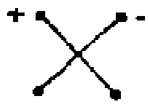
要解这组方程，先将第一个方程乘以 b_2，再将第二个方程乘以 $-b_1$，然后相加，我们就得到

$$(a_1b_2-a_2b_1)x=c_1b_2-c_2b_1.$$

同理，将第一个方程乘以 $-a_2$，将第二个方程乘以 a_1，然后再相加，就得到

$$(a_1b_2-a_2b_1)y=a_1c_2-a_2c_1.$$

只要这两个等式里 x 和 y 的系数 $a_1b_2-a_2b_1$ 不等于零，就容易从这两个等式里定出 x 和 y 来. $a_1b_2-a_2b_1$ 这个式子称为方程组的系数所组成的矩阵 $\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}$ 的**行列式**，行列式记作 $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$. 从上面给出的定义推知，行列式可以按照下图进行计算：



这个图的含义自明，不再需要解释.

现在讨论方程组的解. 根据行列式的定义，$c_1b_2-c_2b_1$ 和 $a_1c_2-a_2c_1$ 这两个式子也可以表作行列式，即 $\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}$ 和 $\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}$.

这样一来，如果行列式 $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$ 不等于零，我们就得到方程组的解的公式：

$$x=\frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}};\quad y=\frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}. \tag{4}$$

严格地说，上面进行的讨论是不完备的. 我们在导出方程组的解的公式时，对于方程所作的运算只在假定了 x 和 y 是组成方程组的解的两个数时才可理解. 上面所进行的讨论的逻辑实质是，如果方程组的系数组成的行列式不等于零而方程组的解存在

的话，那么它的解可以按照公式(4)来计算．因此，还需要证明所求得的未知数的值的确适合方程组里的两个方程．这做起来并没有困难．

这样一来，如果方程组的系数组成的行列式不等于零，那么方程组就有唯一的一组解，而这组解由公式(4)给出．

对于三个未知数的三个方程组

$$a_1x+b_1y+c_1z=d_1,$$
$$a_2x+b_2y+c_2z=d_2,$$
$$a_3x+b_3y+c_3z=d_3$$

不难进行类似的讨论与计算——这需要将三个方程各乘以适当的数之后相加，使得相加以后消去两个未知数．为了消去未知数 y 和 z，应该取 $b_2c_3-b_3c_2$，$b_3c_1-b_1c_3$，$b_1c_2-b_2c_1$ 作为乘数，这是容易验算的．

结果我们得到：如果式子

$$\Delta=a_1b_2c_3-a_1b_3c_2+a_2b_3c_1-a_2b_1c_3+a_3b_1c_2-a_3b_2c_1$$

异于零，那么方程组有唯一解，它可以按公式

$$x=\frac{\Delta_1}{\Delta},\ y=\frac{\Delta_2}{\Delta},\ z=\frac{\Delta_3}{\Delta}$$

求得，其中 Δ_1，Δ_2，Δ_3 是在 Δ 中将相当的未知数的系数用自由项来代替所得的式子．

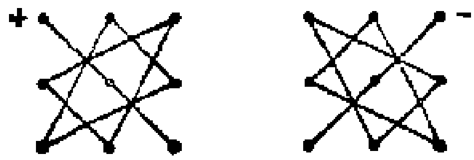
式子 Δ 称为矩阵

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

的行列式，记作

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

为了计算行列式，可参考下面的图：



在第一图里，用线(一个对角线和两个三角形)所联结起来的位置上的元素之积附以正号是行列式的展开式中的项，而在第二个图里，用线所联结起来的位置上的元素之积附以负号是行列式的展开式中的项.

我们对于两个含两个未知数的方程组及三个含三个未知数的方程组得到了完全类似的结果. 在这两种情形，如果系数矩阵的行列式不等于零，方程组就有唯一解. 解的公式也是类似的，每一个未知数的分母都是系数矩阵的行列式，而分子是将系数矩阵中把要求的未知数的系数用自由项来代替所得到的矩阵的行列式.

把这些结果直接推广到含 n 个未知数的 n 个方程组是困难的，用间接的方法来做比较容易做到，即先推广行列式的概念到任意阶的方阵上去，并研究行列式的性质，然后再把它的理论应用到方程组的研究上去.

n 阶行列式 研究行列式的展开式：

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1$$

及

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1b_2c_3 - a_1b_3c_2 + a_2b_3c_1 - a_2b_1c_3 + a_3b_1c_2 - a_3b_2c_1$$

时，我们发现，在每一项里都有每行的一个元素和每列的一个元素出现作为因子，而且所有这种可能的乘积都在行列式的展开式中出现，只是前面附以加号或减号. 这个性质可以作为将行列式的概念推广到任意阶方阵上去的基础. 即，n 阶方阵的行列式，或简单地说，n 阶行列式是方阵中从每一行取一个元素从每一列取一个元素作成的所有乘积按一定的规律冠以正号或负号后的代数

和．引进这个一定的规律相当复杂，我们不准备叙述它．重要的是指出，它是由下面这些行列式的重要基本性质能成立而确定的：

1．将行列式的两行交换，行列式的符号改变．

对于2阶及3阶行列式，这个性质容易用直接计算来验证．在一般情形，它的证明基于我们所没有叙述的符号规律．

行列式有一系列其他的重要性质，这些性质使我们能够有成效地在各种各样的理论及数值计算中采用行列式，不管行列式多么复杂，要知道，不难看出 n 阶行列式含有 $n!$ 项，而每一项又由 n 个因子组成，且这些项前面还按照一个复杂的规律冠以符号．

现在把行列式的基本性质列举出来，但不详细证明．

这些性质中的第一个已经叙述在上面了．

2．将行列式的矩阵转置一下，即用列来代替行而不变其顺序，行列式不变值．

这个性质的证明基于对行列式各项前的符号安排的规律的分析．

3．行列式是它任意一行（或列）的元素的线性函数．详细地说，

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{i1} & \cdots & a_{in} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}, \qquad (5)$$

其中 $A_{i1}, A_{i2}, \cdots, A_{in}$ 是不依赖于 i 行元素的式子．

显然这个性质由行列式的每一项都含某一行的，例如第 i 行的，一个元素而且只含它的一个元素得出．

等式(5)称为将行列式按 i 行元素展开，而 $A_{i1}, A_{i2}, \cdots, A_{in}$ 称为元素 $a_{i1}, a_{i2}, \cdots, a_{in}$ 在行列式中的代数余子式．

4．元素 A_{ij} 的代数余子式除去差一个符号外等于行列式的所谓子式 Δ_{ij}，即从原矩阵划去第 i 行与第 j 列所得的 $(n-1)$ 阶矩

阵的行列式．为了从子式得到代数余子式需要在它前面放置符号 $(-1)^{i+j}$．性质 3 和 4 将 n 阶行列式的计算归结为 $(n-1)$ 阶行列式的计算．

从所列举的这些基本性质推出行列式的许多有用的性质．我们列举出它们中的一些来．

5．具有两个同样的行的行列式等于零．

实际上，如果行列式有两个同样的行，那么将它们对换一下，行列式之值并不改变，因为这两行一样，可是同时根据第一个性质又要改变符号，因此它等于零．

6．任意一行的元素与另外一行元素的代数余子式的乘积之和等于零．

实际上，这样的和是将一个具有两个相同的行的行列式按其中一行展开的结果．

7．任意一行的元素的公因子可以提到行列式符号的外面来．

这由性质 3 推出．

8．含两个成比例的行的行列式之值等于零．

只要把比例因子提出来就得到一个含两个同样的行的行列式．

9．如果将行列式中某一行的元素加上与另一行的元素成比例的数，行列式之值不变．

实际上，根据性质 3，这样造的行列式等于原来的行列式与含两个成比例的行的行列式之和，而后面这个行列式等于零．

最后这个性质给出计算行列式的一个好方法．利用这个性质，可以将行列式的矩阵变成一个某一行(或列)的元素除一个以外都为零的矩阵而不变行列式之值．然后就可以将这个行列式按这一行(或列)展开，我们就把 n 阶行列式的计算化为 $(n-1)$ 阶的一个行列式的计算，即只需计算所选的这一行里唯一的那个不等于零的元素的代数余子式．

例如，需要计算行列式

$$\Delta=\begin{vmatrix} 1 & 1 & -1 & 2 \\ 2 & -1 & 1 & 1 \\ -1 & 2 & 0 & 1 \\ 1 & 1 & -2 & 1 \end{vmatrix}.$$

将第一列乘以 -1 加到第二列，将第一列加到第三列，并将第一列乘以 -2 加到第四列，就得到

$$\Delta=\begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & -3 & 3 & -3 \\ -1 & 3 & -1 & 3 \\ 1 & 0 & -1 & -1 \end{vmatrix}.$$

将 Δ 按第一行的元素展开就得到

$$\Delta=1\cdot(-1)^{1+1}\begin{vmatrix} -3 & 3 & -3 \\ 3 & -1 & 3 \\ 0 & -1 & -1 \end{vmatrix}.$$

最后，将第二行加到第一行，再按第一列展开就得到

$$\Delta=\begin{vmatrix} 0 & 2 & 0 \\ 3 & -1 & 3 \\ 0 & -1 & -1 \end{vmatrix}=3\cdot(-1)^{1+2}\begin{vmatrix} 2 & 0 \\ -1 & -1 \end{vmatrix}=-3\cdot(-2)=6.$$

方阵 A 的行列式记作 $|A|$.

最后我们再举出行列式的一个重要性质.

两个方阵的乘积的行列式等于这两个方阵的行列式的乘积，即 $|AB|=|A|\cdot|B|$.

特别，这个性质使我们能将同阶的行列式按矩阵乘法规律相乘.

含 n 个未知数的 n 个线性方程组 利用行列式很容易将前面得到的关于含两个未知数的两个方程组及含三个未知数的三个方程组的结果推广到含 n 个未知数的 n 个方程组来，只要方程组的系数矩阵的行列式不等于零.

设 $$a_{11}x_1+\cdots+a_{1j}x_j+\cdots+a_{1n}x_n=b_1,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_{n1}x_1+\cdots+a_{nj}x_j+\cdots+a_{nn}x_n=b_n,$$

是这样一个方程组,以 Δ 表这个方程组的系数矩阵的行列式,再以 A_{ij} 表元素 a_{ij} 的代数余子式. 将第一个方程乘以 A_{1j},第二个方程乘以 A_{2j}, …, 第 n 个方程乘以 A_{nj}, 然后相加,我们得到

$$\Delta x_j=b_1A_{1j}+\cdots+b_nA_{nj}.$$

实际上,除了 x_j 的系数之外,其余未知数的系数都等于零,因为它们是 j 列元素的代数余子式与另一列元素的乘积的和(性质6, 运用到列上); 而 x_j 的系数等于 j 列元素与它的代数余子式的乘积的和,因而等于 Δ.

这样一来,

$$x_j=\frac{b_1A_{1j}+\cdots+b_nA_{nj}}{\Delta}\text{ 对一切 }j=1,\ 2,\ \cdots,\ n. \tag{6}$$

像前面所说的一样,上面所进行的论述只在假定了方程组解的存在性并将 x_1, x_2, …, x_n 了解为方程组的解时才有意义.

因此,上面论述的结论如下:

如果方程组的解存在,那么它由公式(6)给出而且是唯一的.

为了论述的完整,还需要证明方程组的解的存在,为此需将所得值代替未知数代入原方程组的一切方程. 采用行列式的同一性质(应用到行)容易验证,求得的值的确适合一切方程.

于是,下面的定理成立:如果含 n 个未知数的 n 个方程组的系数矩阵的行列式不等于零,那么这个方程组有唯一解,它的解由公式(6)给出.

这个公式还可以变形,即和 $b_1A_{1j}+\cdots+b_nA_{nj}$ 可以写作行列式的形状:

$$\Delta_j=b_1A_{1j}+\cdots+b_nA_{nj}=\begin{vmatrix} a_{11} & \cdots & b_1 & \cdots & a_{1n}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ a_{n1} & \cdots & b_n & \cdots & a_{nn}\end{vmatrix}$$

(自由项位于第 j 列).

这样一来,前面对于含两个未知数的方程组及含三个未知数的方程组所述的结果完全推广到含 n 个未知数的方程组,而所得

到的解的公式也是有同样形状的.

我们注意上面所证明的定理的一个推论.

如果已经知道方程组根本没有解或解不唯一，那么它的系数矩阵的行列式等于零.

这个推论特别可以应用到齐次方程组，即自由项 b_1, b_2, …, b_n 都等于零的方程组. 齐次方程组总有所谓"显然"解 $x_1=x_2=\cdots=x_n=0$.

如果齐次方程组除了显然解之外还有非显然解，那么它的行列式等于零.

这个断言打开了在其他数学部门及其应用中采用行列式论的可能性.

例如，考虑解析几何的一个问题.

求通过不在同一直线上的三个已知点 (x_1, y_1, z_1), (x_2, y_2, z_2) 和 (x_3, y_3, z_3) 的平面的方程.

从初等几何知道，所求的平面是存在的. 设它的方程是 $Ax+By+Cz+D=0$, 那么

$$Ax_1+By_1+Cz_1+D=0,$$
$$Ax_2+By_2+Cz_2+D=0,$$
$$Ax_3+By_3+Cz_3+D=0.$$

设 x, y, z 是这个平面上任意点的坐标，那么还有

$$Ax+By+Cz+D=0.$$

将这四个方程看作是对于欲求平面的系数 A, B, C, D 的齐次线性方程组. 这个方程组有非显然的解，因为欲求的平面存在. 因此，方程组的行列式等于零，即

$$\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x & y & z & 1 \end{vmatrix}=0. \tag{7}$$

这就是欲求平面的方程. 实际上，它是 x, y, z 的一次方程，这从行列式对于最后一行元素的线性性质即可推出.

采用已知的三个点不在同一直线上这一事实，不难验证这个方程的系数不可能都等于零．因此，等式(7)的确是一个平面的方程．这个平面通过已知点，因为它们的坐标显然适合方程．

含 n 个未知数的 n 个方程的矩阵写法 含 n 个未知数的 n 个线性方程组

$$a_{11}x_1+\cdots+a_{1n}x_n=b_1,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$a_{n1}x_1+\cdots+a_{nn}x_n=b_n$$

可以用矩阵符号写成一个等式的形状：

$$AX=B.$$

这里 A 表系数矩阵，X 表由未知数所组成的列，B 表自由项所组成的列．

方程组的解(如果矩阵 A 的行列式不等于零)可以详细地写作(参看公式(6))：

$$x_1=\frac{A_{11}}{\Delta}b_1+\frac{A_{21}}{\Delta}b_2+\cdots+\frac{A_{n1}}{\Delta}b_n,$$
$$x_2=\frac{A_{12}}{\Delta}b_1+\frac{A_{22}}{\Delta}b_2+\cdots+\frac{A_{n2}}{\Delta}b_n,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$x_n=\frac{A_{1n}}{\Delta}b_1+\frac{A_{2n}}{\Delta}b_2+\cdots+\frac{A_{nn}}{\Delta}b_n$$

或写成矩阵形式

$$X=\begin{pmatrix}\frac{A_{11}}{\Delta} & \frac{A_{21}}{\Delta} & \cdots & \frac{A_{n1}}{\Delta}\\ \frac{A_{12}}{\Delta} & \frac{A_{22}}{\Delta} & \cdots & \frac{A_{n2}}{\Delta}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{A_{1n}}{\Delta} & \frac{A_{2n}}{\Delta} & \cdots & \frac{A_{nn}}{\Delta}\end{pmatrix}B.$$

作为等式右方第一个因子的矩阵称为 A 的**逆矩阵**，记作 A^{-1}．采用了这个记号，我们就得到方程组 $AX=B$ 解的下面这个简单而自然的形状：

$$X=A^{-1}B,$$

它好像含一个未知数的一个方程的解的公式.

不难用矩阵代数的语言,给上面所获得的结果以另外的论据.

为了这个目的,首先应该指出起着特殊作用的矩阵

$$E=\begin{pmatrix}1 & 0 & 0 & \cdots & 0\\ 0 & 1 & 0 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & 0 & \cdots & 1\end{pmatrix}.$$

它称为**单位矩阵**.

单位矩阵在方阵中起着1这个数在全体数中所起的那样的作用,即,对于任何矩阵 A,等式 $AE=A$ 和 $EA=A$ 都成立. 根据矩阵乘法的规律很容易验证这一点.

上面所定义的矩阵 A 的逆矩阵 A^{-1} 所起的作用,类似于一个数的倒数所起的作用,即

$$AA^{-1}=A^{-1}A=E.$$

可以根据矩阵的乘法规律与行列式的性质3和6来验证这些等式的真实性.

知道了单位矩阵和逆矩阵的这些性质之后, 可以按照下列方式求方程组 $AX=B$ 的解.

设 $AX=B$,则 $A^{-1}(AX)=A^{-1}B$. 可是

$$A^{-1}(AX)=(A^{-1}A)X=EX=X,$$

因此 $X=A^{-1}B$.

现在设 $X=A^{-1}B$, 那么 $AX=AA^{-1}B=EB=B$.

所以,"方程" $AX=B$ 有唯一解 $X=A^{-1}B$, 只要 A^{-1} 存在.

我们在矩阵 A 的行列式不等于零的假设下,证明了 A 的逆矩阵 A^{-1} 存在. 这个条件对于逆矩阵的存在不仅充分而且必要. 实际上, 设对于矩阵 A 存在着逆矩阵 A^{-1}, 即使得 $AA^{-1}=E$. 那么根据两个矩阵乘积的行列式的性质,就有

$$|A|\cdot|A^{-1}|=|E|=1.$$

由此推出,矩阵 A 的行列式不等于零.

其行列式不等于零的矩阵称为非退化的或非奇异的. 这样,我们证明了对于非退化矩阵,逆矩阵永远存在,而且只对于它们逆矩阵才存在.

逆矩阵概念的引进不仅对于线性方程组的理论有用, 对线性代数的其他一些问题也有用.

最后我们指出, 上面引进的解线性方程组的公式是理论性的研究中不可缺少的工具, 但在方程组的数值解法中却很少被采用.

正象我们已经指出的, 方程组的数值解法有很多各式各样的方法和计算方案已经求得. 由于这个问题对于实际的重要性, 使得方程组(特别是大量未知数的情形)数值解法的简化方面的研究工作, 目前正在紧张地进行着.

线性方程组的一般情形 现在我们转到最一般情形的线性方程组的研究, 甚至不假定方程的个数等于未知数的个数. 在这个一般情形, 自然不能期望方程组的解永远存在或当它存在时是唯一的. 自然会预料到, 如果方程的个数少于未知数的个数, 方程组会有无限多组解. 例如, 含三个未知数的两个一次方程被这两个一次方程所定义的平面的交线上的一切点所适合. 当然也可能在这个情形方程组根本没有解, 即当两张平面平行时就如此. 如果方程的个数多于未知数的个数, 一般说来是无解的. 然而在这种情形下, 方程组也可能有解, 甚至有无限多组解.

为了研究一般情形的方程组的解的存在, 以及由方程组的解所成集合的性质, 我们现在讨论方程组的"几何"解释.

方程组

$$\begin{aligned} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n&=b_1,\\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n&=b_2,\\ &\cdots\cdots\\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n&=b_m \end{aligned} \tag{8}$$

在 m 维列空间中可解释作

$$x_1\boldsymbol{A}_1+x_2\boldsymbol{A}_2+\cdots+x_n\boldsymbol{A}_n=\boldsymbol{B}.$$

这里 $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n$ 表示由相当的未知数前的系数所组成的列，而 $\boldsymbol{B}$ 表示由自由项所组成的列.

在这个解释下，方程组解的存在问题就化为已知向量 $\boldsymbol{B}$ 是否可表作向量 $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n$ 的线性组合这个问题.

关于这个问题的解答几乎是显然的. 向量 $\boldsymbol{B}$ 是向量 $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n$ 的线性组合的充分必要条件是向量 $\boldsymbol{B}$ 属于向量 $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n$ 所张成的向量子空间，换句话说，向量组 $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n$ 所张成的子空间与向量组 $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n, \boldsymbol{B}$ 所张成的子空间相重.

因为第一个子空间包在第二个子空间之中，那么它们相重的充分必要条件是它们的维数相等. 我们已经讲过，已知向量组所张成的子空间的维数称为这组向量的秩. 这样一来，方程组 $x_1\boldsymbol{A}_1+x_2\boldsymbol{A}_2+\cdots+x_n\boldsymbol{A}_n=\boldsymbol{B}$ 解存在的充分必要条件就是向量组 $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n$ 的秩与向量组 $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n, \boldsymbol{B}$ 的秩相等.

可以证明，向量组的秩等于由这些向量的坐标所组成的矩阵的秩. 我们不拟证明. 这里我们将**矩阵的秩**理解为从矩阵划去一部分行及列所能组成的不等于零的行列式的最大阶数.

因为方程组的系数是向量 $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n$ 的坐标（对于列空间的自然基而言），而它的自由项是 $\boldsymbol{B}$ 的坐标，所以我们得到下面这个彻底解决了的关于方程组解存在条件的结论.

线性方程组

$$
\begin{gathered}
a_{11}x_1+\cdots+a_{1n}x_n=b_1,\\
\cdots\cdots\cdots\cdots\cdots\cdots\\
a_{m1}x_1+\cdots+a_{mn}x_n=b_m
\end{gathered}
$$

至少存在一组解的充分必要条件是由方程组的系数所组成的矩阵的秩等于由方程组的系数及自由项所组成的矩阵的秩.

现在来研究方程组的解所成的集合的性质. 如果解存在，设方程组 (8) 的任意一组解是 $x_1^0, x_2^0, \cdots, x_n^0$. 令 $x_1=x_1^0+y_1$, $x_2=x_2^0+y_2, \cdots, x_n=x_n^0+y_n$. 这时因为 $x_1^0, x_2^0, \cdots, x_n^0$ 组成(8)的一组解，所以未知数 $y_1, y_2, \cdots, y_n$ 需要适合齐次方程组

$$
\begin{aligned}
&a_{11}y_1+\cdots+a_{1n}y_n=0,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&a_{m1}y_1+\cdots+a_{mn}y_n=0,
\end{aligned}
\tag{9}
$$

其系数矩阵与(8)的相同. 反之，将(8)的一组解 x_1^0, x_2^0, $\cdots$, x_n^0 加上齐次方程组(9)的任意一组解，仍然得到(8)的一组解.

因此，为了获得方程组(8)的通解，需要取它的任意一个特解，然后再加上方程组(9)的通解.

这样一来，关于方程组(8)的解的集合性质的问题就化为齐次方程组(9)的同一问题了. 这个问题在下面讨论.

齐次方程组 我们要在 n 维欧氏空间中来解释齐次线性方程组

$$
\begin{aligned}
&a_{11}y_1+a_{12}y_2+\cdots+a_{1n}y_n=0,\\
&a_{21}y_1+a_{22}y_2+\cdots+a_{2n}y_n=0,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&a_{m1}y_1+a_{m2}y_2+\cdots+a_{mn}y_n=0.
\end{aligned}
$$

(这样我们要假定方程组的系数是实数. 对于复系数的方程可以在酉空间中给出类似的解释，因而得出类似的结果.)

设 $\boldsymbol{A}_1'$, $\boldsymbol{A}_2'$, $\cdots$, $\boldsymbol{A}_m'$, $\boldsymbol{Y}$ 是欧氏空间中的向量，它们在某个标准正交基下的坐标分别是

$$
(a_{11},\ a_{12},\ \cdots,\ a_{1n}),\ (a_{21},\ a_{22},\ \cdots,\ a_{2n}),\ \cdots,
$$
$$
(a_{m1},\ a_{m2},\ \cdots,\ a_{mn}),\ (y_1,\ y_2,\ \cdots,\ y_n).
$$

这时方程组采取下列形状：

$$
\boldsymbol{A}_1'\boldsymbol{Y}=0,\ \boldsymbol{A}_2'\boldsymbol{Y}=0,\ \cdots,\ \boldsymbol{A}_m'\boldsymbol{Y}=0,
$$

即方程组的每一个解都确定一个与诸方程的系数所组成的向量都正交的向量.

因此，解的集合组成向量 $\boldsymbol{A}_1'$, $\boldsymbol{A}_2'$, $\cdots$, $\boldsymbol{A}_m'$ 所张成的子空间的正交补空间. 前面这个子空间的维数等于方程组的系数矩阵的秩 r. 于是它的正交补空间，即"解空间"的维数等于 $n-r$.

在每一个子空间里都存在着一组基，即一组线性无关的向量，它的个数等于子空间的维数，而子空间中任何一个向量都可表作

它们的线性组合．因此，在齐次方程组的解里存在着 $n-r$ 个线性无关的解，使得方程组的一切解都是它们的线性组合．这里 n 表示未知数的个数，r 表示系数矩阵的秩．

于是，齐次方程组解的构造，以及非齐次方程组解的构造，就完全阐述清楚了．特别，齐次方程组当且仅当系数矩阵的秩等于未知数的个数时，以显然解 $x_1=x_2=\cdots=x_n=0$ 为它的唯一解．基于上段最后所述的结果，这个条件(当相容性条件成立时)也是非齐次方程组解的唯一性的条件．

我们所进行的方程组的研究明显的示出了，推广的几何概念的引进如何给复杂的代数问题带来简单与清楚．

§4．线性变换

定义和例子 在很多数学研究中需要改换变数，即从一组变数 $x_1, x_2, \cdots, x_n$ 过渡到与它们有函数关系的另一组变数 $y_1, y_2, \cdots, y_n$:

$$
\begin{aligned}
y_1&=\varphi_1(x_1, x_2, \cdots, x_n),\\
y_2&=\varphi_2(x_1, x_2, \cdots, x_n),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
y_n&=\varphi_n(x_1, x_2, \cdots, x_n).
\end{aligned}
$$

例如，如果变数是平面上或空间中点的坐标，那么从一个坐标系过渡到另一个坐标系就引起坐标的一个变换，它将原来的坐标用新的坐标表出，或反之．

此外，在研究一个物体从一个位置或状态变为另一个位置或状态时，如果它的位置或状态由变数的值所给出，变数的变换也会产生．这种变换的典型例子就是某物体在变形之下它的点的坐标的改变．

抽象给出的一组 n 个变数的变换通常解释作 n 维空间的变换(变形)，即将空间中(或它一部分里)具有坐标 $x_1, x_2, \cdots, x_n$ 的向量用具有坐标 $y_1, y_2, \cdots, y_n$ 的向量来代替．

上面已经说过，每一个多变数的“光滑”(有连续偏导数的)函数，在这些变数改动得很小的时候，与线性函数很接近. 因此任何“光滑”变换(即，在它的解析表达式中的函数 $\varphi_1, \varphi_2, \cdots, \varphi_n$ 有连续偏导数)，在小范围内，与线性变换

$$\begin{aligned}
y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + b_1,\\
y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + b_2,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
y_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n + b_n
\end{aligned} \tag{10}$$

很接近.

这种情况说明线性变换的性质的研究是数学中的重要问题之一. 例如，从含 n 个未知数的 n 个线性方程的理论中我们知道，对于 $x_1, x_2, \cdots, x_n$ 的方程组(10)的解唯一存在，也即相应的线性变换可逆的充分必要条件是它的系数行列式不等于零. 这个情况是分析中一个深刻定理的基础：在已知点光滑的变换

$$\begin{aligned}
y_1 &= \varphi_1(x_1, x_2, \cdots, x_n),\\
y_2 &= \varphi_2(x_1, x_2, \cdots, x_n),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
y_n &= \varphi_n(x_1, x_2, \cdots, x_n),
\end{aligned}$$

有光滑的逆变换的条件是在这一点行列式

$$\begin{vmatrix}
\dfrac{\partial\varphi_1}{\partial x_1} & \cdots & \dfrac{\partial\varphi_1}{\partial x_n}\\
\dfrac{\partial\varphi_2}{\partial x_1} & \cdots & \dfrac{\partial\varphi_2}{\partial x_n}\\
\cdots & \cdots & \cdots\\
\dfrac{\partial\varphi_n}{\partial x_1} & \cdots & \dfrac{\partial\varphi_n}{\partial x_n}
\end{vmatrix}$$

不等于零.

一般的线性变换(10)的研究，基本上可以化为具有同样系数的齐次线性变换

$$\begin{aligned}
y_1 &= a_{11}x_1 + \cdots + a_{1n}x_n,\\
y_2 &= a_{21}x_1 + \cdots + a_{2n}x_n,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
y_n &= a_{n1}x_1 + \cdots + a_{nn}x_n
\end{aligned} \tag{11}$$

的研究；今后凡谈及线性变换，我们永远指齐次变换.

n 维空间中的线性变换也可以用它的内在性质来定义，而不用联系着对应点的坐标的公式(11)来定义. 这种不用坐标的线性变换的定义，好处在于它不依赖于基底的选取. 这个定义如下：

n 维线性空间中的线性变换是一个函数 $\boldsymbol{Y}=\boldsymbol{A}(\boldsymbol{X})$，其中变数 $\boldsymbol{X}$ 与函数值 $\boldsymbol{Y}$ 都是向量. 这个函数是线性的：

$$\boldsymbol{A}(c_1\boldsymbol{X}_1+c_2\boldsymbol{X}_2)=c_1\boldsymbol{A}(\boldsymbol{X}_1)+c_2\boldsymbol{A}(\boldsymbol{X}_2). \qquad (12)$$

今后谈到空间的线性变换时，我们永远按照这个定义的意义来理解.

这个定义和上面利用坐标给出的定义等价. 实际上，如将坐标为 $x_1, x_2, \cdots, x_n$ 的向量 $\boldsymbol{X}$ 代之以坐标为 $y_1, y_2, \cdots, y_n$ 的向量 $\boldsymbol{Y}$，使得坐标 $y_1, y_2, \cdots, y_n$ 表成坐标 $x_1, x_2, \cdots, x_n$ 的线性齐次函数，那么函数 $\boldsymbol{Y}=\boldsymbol{A}(\boldsymbol{X})$ 显然适合条件 (12). 反之，如果函数 $\boldsymbol{Y}=\boldsymbol{A}(\boldsymbol{X})$ 适合条件 (12)，而 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 是空间的任意一组基，那么

$$\boldsymbol{A}(x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+\cdots+x_n\boldsymbol{e}_n)=x_1\boldsymbol{A}(\boldsymbol{e}_1)+x_2\boldsymbol{A}(\boldsymbol{e}_2)+\cdots+x_n\boldsymbol{A}(\boldsymbol{e}_n).$$

将向量 $\boldsymbol{A}(\boldsymbol{e}_j)$ 的坐标(在同一组基之下)记作 $a_{1j}, \cdots, a_{nj}$; $j=1, \cdots, n$, 那么向量 $\boldsymbol{Y}=\boldsymbol{A}(\boldsymbol{X})$ 的坐标就是

$$\begin{aligned} y_1 &= a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n, \\ y_2 &= a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n &= a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n. \end{aligned}$$

这样一来，线性空间中每一个线性变换都对应着一个方阵. 变换本身可以用矩阵语言写成形状 $Y=AX$，这里 X 是原向量的坐标组成的列，Y 是变换后的向量的坐标组成的列，A 是变换的系数矩阵. 矩阵 A 的列由基向量所变成的向量的坐标所组成. 与矩阵写法相应，今后我们常把线性变换本身写作形状 $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}$，而略去括号.

由公式

$$\boldsymbol{A}(x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+\cdots x_n\boldsymbol{e}_n)=x_1\boldsymbol{A}(\boldsymbol{e}_1)+x_2\boldsymbol{A}(\boldsymbol{e}_2)+\cdots+x_n\boldsymbol{A}(\boldsymbol{e}_n)$$

推出，整个空间在线性变换之下变到了由向量 $\boldsymbol{A}(\boldsymbol{e}_1)$, $\boldsymbol{A}(\boldsymbol{e}_2)$, …, $\boldsymbol{A}(\boldsymbol{e}_n)$ 所张成的子空间. 这个子空间的维数等于向量组 $\boldsymbol{A}(\boldsymbol{e}_1)$, $\boldsymbol{A}(\boldsymbol{e}_2)$, …, $\boldsymbol{A}(\boldsymbol{e}_n)$ 的秩，或者说等于它们的坐标所组成的矩阵的秩，即与变换相应的矩阵 A 的秩. 这个子空间，当且仅当矩阵 A 的秩等于 n，即当矩阵 A 的行列式不等于零时，与整个空间相重，这时线性变换称为非奇异的或非退化的.

从线性方程组的理论我们知道，非退化的矩阵有唯一的逆矩阵，同时原来向量的坐标可以用变换后的向量的坐标以公式 $X=A^{-1}Y$ 表出.

矩阵的行列式之值等于零的变换称为奇异的或退化的. 退化变换是不可逆的. 这可以从线性方程的理论推出，或较直观地从它将整个空间映到它的一部分这一事实推出.

将所有向量都映到它自身的单位变换是非退化变换中首先要提到的例子. 单位变换在任何一组基之中的矩阵都是单位矩阵 E.

将空间中所有向量都乘以同一组不等于零的数的相似变换也是非退化的. 相似变换的矩阵不依赖于基的选取，它的形状是 aE，这里 a 是相似系数.

正交变换是非退化变换的重要特殊情形. 正交变换的概念在应用于欧氏空间时才有意义，它定义为将向量长度保持不变的线性变换. 正交变换是将空间中坐标原点不动的旋转或旋转与对通过原点的某一平面的反射的联合对 n 维空间的推广.

显而易见，正交变换不仅保持向量的长度不变，也保持内积不变. 因此，正交变换将空间的正交基变到一组两两正交的单位向量. 它们当然也是一组正交基.

在正交基之下，与正交变换相应的矩阵有着下面这些特殊性质.

首先，任何一列元素的平方和等于1，因为这样的和是由选定的基中的向量所变换到的向量的长度的平方. 其次，两个不同的列的相应元素之积的和等于零，因为这样的和是两个基向量所变

换到的两个向量的数量积.

采用矩阵记号,这两个性质可以用一个公式写出:

$$\bar{P}P=E.$$

这里 P 是正交变换的矩阵(对于正交基), $\bar{P}$ 是 P 的转置矩阵,即它的行是矩阵 P 的列并保持原来顺序.

实际上,根据矩阵乘法规律,矩阵 $\bar{P}P$ 的对角元素等于矩阵 P 中相应列的元素的平行和,而非对角元素等于取自矩阵 P 中不同列的相应元素之积的和.

退化变换的例子有欧氏空间中所有向量在某个子空间上的正交射影(参看 §2). 实际上,在这个变换下,整个空间映为它自己的一部分.

坐标变换 现在来研究 n 维空间中坐标变换的问题,即从一组基转变到另一组基时,向量的坐标如何改变的问题.

设原来的基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 已给定,并设 $\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_n$ 是空间的任意另一组基. 再设 $C=\begin{pmatrix} c_{11}\cdots, & c_{1n} \\ \cdots\cdots & \cdots \\ c_{n1}\cdots, & c_{nn}\end{pmatrix}$ 是一个矩阵,它的列是新基中的向量 $\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_n$ 对于原来的基的坐标. 由于向量 $\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_n$ 的线性无关性,显然矩阵 C 是非退化的. 它称为坐标变换矩阵.

以 $x_1, x_2, \cdots, x_n$ 表示向量 $\boldsymbol{X}$ 对于基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 的坐标,以 $x'_1, x'_2, \cdots, x'_n$ 表示它对于基 $\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_n$ 的坐标,那么 $\boldsymbol{X}=x'_1\boldsymbol{f}_1+x'_2\boldsymbol{f}_2+\cdots+x'_n\boldsymbol{f}_n$. 因之向量 $\boldsymbol{X}$ 对于原来的基的坐标组成的列

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} c_{11}x'_1+c_{12}x'_2+\cdots+c_{1n}x'_n \\ c_{21}x'_1+c_{22}x'_2+\cdots+c_{2n}x'_n \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ c_{n1}x'_1+c_{n2}x'_2+\cdots+c_{nn}x'_n \end{pmatrix}=\begin{pmatrix} c_{11} & \cdots & c_{1n} \\ c_{21} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}.$$

于是,原来的坐标可以用变换后的坐标以矩阵 C 齐次线性地表出.

表明对于原来基的坐标与对于变换后的基的坐标之间的依赖关系的公式，与向量在空间的非退化线性变换下坐标之间关系的公式形式上相同。这种情况使我们能够将抽象给出的变数的具有非退化矩阵的线性齐次变换，或者解释作坐标的变换，或者解释作空间的线性变换。在每一个具体场合，这两种解释的选择由考虑的问题的内容所决定。

现在来研究空间的线性变换的矩阵在坐标变换下改动的情况。

设对于基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 而言，线性变换的矩阵是 A，即变换后的向量的坐标所组成的列 Y 与原来的向量的坐标所组成的列 X 由公式

$$Y = AX$$

相联系。

设现在做好了一个具有矩阵 C 的坐标变换；X'，Y' 分别表示原来的向量和变换后的向量对于新基的坐标。这时 $X = CX'$，$Y = CY'$。由此推出

$$Y' = C^{-1}Y = C^{-1}AX = C^{-1}ACX'.$$

这样，所考虑的变换对于新基的矩阵就是 $C^{-1}AC$。

矩阵 A 和 B 如果由公式 $B = C^{-1}AC$ 联系起来，其中 C 为某个非奇异矩阵，称为**相似**。在不同基之下对应于同一个线性变换的矩阵，组成两两相似的一组矩阵。

线性变换的特征向量和特征值　一类重要的线性变换由下面这种方法构成：

设 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 是空间中任意一组线性无关的向量。设在一个变换下它们分别乘以 $\lambda_1, \lambda_2, \cdots, \lambda_n$。如果向量 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 取作空间的基，那么所考虑的线性变换可用对角矩阵

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

来表出.

这类变换有着简单而直观的几何意义(当然是指对实空间且 $n=2$ 和 $n=3$ 时而言),即,如果 λ_i 这些数都是正数,那么这个变换就将空间按向量 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 的方向各伸展 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 倍. 如果某些 λ_i 是负的,那么空间变形的同时,向量 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 中某些方向将改为反方向. 最后,例如当 $\lambda_1=0$ 时,要先将空间沿着平行于 $\boldsymbol{e}_1$ 的方向投影到由 $\boldsymbol{e}_2, \boldsymbol{e}_3, \cdots, \boldsymbol{e}_n$ 所张成的子空间上,然后再按着这些方向作变形.

上面所考虑的这类变换的重要性在于,尽管它是简单的,但却是很一般的. 可以证明,适合某种不太强的限制的线性变换都属于这类变换,即可以选取一组基,使这个变换用对角矩阵表出.

如果研究复空间中线性变换的话,加在变换上的限制将变得特别弱.今后我们就在复空间上讨论.

现在引进下面的定义.

非零向量 $\boldsymbol{X}$,如在线性变换 $\boldsymbol{A}$ 之下变为与它共线的向量 $\lambda\boldsymbol{X}$,就称为变换的特征向量. 换言之,非零向量 $\boldsymbol{X}$ 当且仅当 $\boldsymbol{AX}=\lambda\boldsymbol{X}$ 时是变换 $\boldsymbol{A}$ 的特征向量. λ 这个数称为变换 $\boldsymbol{A}$ 的特征值.

显然,如果变换在某个基之下的矩阵是对角矩阵,那么这个基由特征向量组成,而矩阵的对角元素就是特征值. 反之,如果空间中存在着由变换 $\boldsymbol{A}$ 的特征向量所组成的一组基,那么在这个基之下变换 $\boldsymbol{A}$ 的矩阵是对角矩阵,而且其对角线由对应这组基中向量的特征值组成.

现在来研究特征向量和特征值的性质. 为了这个目的,我们在特征向量的定义中给出坐标写法. 设 A 是变换 $\boldsymbol{A}$ 对应于某个基的矩阵, X 是由向量 $\boldsymbol{X}$ 在这个基里的坐标组成的列. 将等式 $\boldsymbol{AX}=\lambda\boldsymbol{X}$ 转化为坐标形式,就写成 $AX=\lambda X$ 或

$$(A-\lambda E)X=0.$$

展开之后,这个等式就化为方程组

$$(a_{11}-\lambda)x_1+a_{12}x_2+\cdots+a_{1n}x_n=0,$$

$$a_{21}x_1+(a_{22}-\lambda)x_2+\cdots+a_{2n}x_n=0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_{n1}x_1+a_{n2}x_2+\cdots+(a_{nn}-\lambda)x_n=0.$$

这个方程组可以看作是关于 $x_1, x_2, \cdots, x_n$ 的齐次线性方程组. 我们有兴趣的情形是这个方程组有非显然解，因为特征向量的坐标不应该同时都等于零. 我们知道，这个方程组存在着非显然解的充分必要条件是系数矩阵的秩小于未知数的个数，而这等价于方程组的行列式等于零

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix}=0.$$

因此，变换 $\boldsymbol{A}$ 的特征值是多项式 $|A-\lambda E|$ 的根，反之，这个多项式的每一个根都是变换的特征值，因为每一个根都至少对应着一个特征向量. 多项式 $|A-\lambda E|$ 称为矩阵 A 的**特征多项式**. 方程 $|A-\lambda E|=0$ 称为矩阵 A 的**特征方程**或**永年方程**[1]，而它的根称为矩阵的**特征值**.

根据高等代数基本定理(第一卷，第四章)，每一个多项式至少有一个根，因此每一个线性变换至少有一个特征值，因而至少有一个特征向量. 可是，即使这个线性变换的矩阵是实矩阵，当然可能它的特征值的全体或一部分是复数. 实际上，在实空间中，对于任意线性变换，(实)特征值及特征向量存在的定理并不成立. 例如，平面上围绕坐标原点转动一个异于 180° 的角这个变换改变了平面上所有向量的方向，因此这个变换的特征值不存在.

矩阵 A 的特征多项式的根是变换 $\boldsymbol{A}$ 的特征值，因此同一变换在不同基之下的矩阵的特征多项式的根的集合相重. 这个事实说明下面这个命题的真实性：线性变换的特征多项式只依赖于变换而不依赖于基底的选择. 这个命题可以用基于矩阵及行列式运算的性质的计算来如下验证：

1) “永年方程”的名称源自天体力学，它与行星运动中所谓的永年扰动问题有关.

我们知道，如果变换 $\boldsymbol{A}$ 在某个基之下的矩阵是 A，那么在另一个基之下的矩阵就是 $C^{-1}AC$，这里 C 是某个非奇异矩阵. 可是

$$|C^{-1}AC-\lambda E|=|C^{-1}AC-C^{-1}\lambda EC|=|C^{-1}(A-\lambda E)C|$$
$$=|C^{-1}||C||A-\lambda E|=|C^{-1}C||A-\lambda E|=|A-\lambda E|.$$

因此，线性变换 $\boldsymbol{A}$ 在不同基底下所对应的矩阵实际上有相同的特征多项式，因此，可以算之为线性变换的特征多项式.

现在假设线性变换 $\boldsymbol{A}$ 的所有特征值都不同，我们来证明：如果对于每一个特征值取一个特征向量，这些特征向量线性无关. 实际上，假定它们中的某几个，例如 $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_k$，线性无关，而其余的，其中有 $\boldsymbol{e}_{k+1}$，是它们的线性组合，那么

$$\boldsymbol{e}_{k+1}=c_1\boldsymbol{e}_1+c_2\boldsymbol{e}_2+\cdots+c_k\boldsymbol{e}_k. \tag{13}$$

将线性变换 $\boldsymbol{A}$ 作用到上式双方，就得到

$$\boldsymbol{A}\boldsymbol{e}_{k+1}=c_1\boldsymbol{A}\boldsymbol{e}_1+c_2\boldsymbol{A}\boldsymbol{e}_2+\cdots+c_k\boldsymbol{A}\boldsymbol{e}_k.$$

由此根据特征向量的定义推出

$$\lambda_{k+1}\boldsymbol{e}_{k+1}=c_1\lambda_1\boldsymbol{e}_1+c_2\lambda_2\boldsymbol{e}_2+\cdots+c_k\lambda_k\boldsymbol{e}_k.$$

将等式(13)乘以 λ_{k+1}，再从它减去上式，就有

$$c_1(\lambda_{k+1}-\lambda_1)\boldsymbol{e}_1+c_2(\lambda_{k+1}-\lambda_2)\boldsymbol{e}_2+\cdots+c_k(\lambda_{k+1}-\lambda_k)\boldsymbol{e}_k=0.$$

根据 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_k$ 的线性无关性，由上式推出

$$c_1(\lambda_{k+1}-\lambda_1)=c_2(\lambda_{k+1}-\lambda_2)=\cdots=c_k(\lambda_{k+1}-\lambda_k)=0.$$

但是我们假定所有的特征值都不同，而且对于每一个特征值选取一个特征向量. 因此 $\lambda_{k+1}-\lambda\neq 0$, $\lambda_{k+1}-\lambda_2\neq 0$, $\cdots$, $\lambda_{k+1}-\lambda_k\neq 0$，所以等式 (13) 不可能成立，因为系数 $c_1, c_2, \cdots, c_k$ 不能同时等于零.

现在清楚了，如果线性变换的所有特征值都不同，那么存在着一组基，在这组基之下变换的矩阵是对角形的. 实际上，对于每一个特征值选定一个特征向量，这些向量的全体就可以选取作为这样的基. 我们已经证明了它们是线性无关的，它们的个数又等于空间的维数，即它们的确组成一组基.

上面证明的定理可以用矩阵论的语言叙述如下.

如果一个矩阵的所有特征值彼此不相等，那么这个矩阵相似

于一个对角矩阵，它的对角线元素就是这些特征值.

将线性变换的矩阵化为最简单的形状这个问题，当特征多项式的根有相等的时候，相当复杂．我们只简单地叙述一下最后的结果.

m 阶的"标准块"是指形状

$$I_{m,\lambda_i}=\begin{pmatrix}\lambda_i & 1 & & \\ & \lambda_i & 1 & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i\end{pmatrix}$$

的矩阵，其中所有没写出来的元素都等于零.

一个矩阵，如果沿着它的主对角线排列着一些"标准块"而其余的元素都等于零

$$\begin{pmatrix}\boxed{I_{m_1,\lambda_1}} & & & \\ & \boxed{I_{m_2,\lambda_2}} & & \\ & & \ddots & \\ & & & \boxed{I_{m_k,\lambda_k}}\end{pmatrix},$$

则称为约当标准矩阵.

不同"块"里的 λ_i 这些数并不一定两两不同．任意矩阵都可以化为与它相似的约当标准矩阵．这个定理的证明相当复杂．但必须指出，这个定理在代数对其他数学问题的应用中，特别在线性微分方程组的理论中，起着很大的作用.

矩阵当且仅当所有块的阶 m_i 都等于 1 时可以化为对角形.

§5. 二 次 型

定义和简单性质 多变数的二次齐次多项式称为**二次型**.

n 个变数 $x_1, x_2, \cdots, x_n$ 的二次型由两种项组成：变数的平方及它们两两的乘积再添上系数．二次型可以写成下面的形状：

$$f(x_1, x_2, \cdots, x_n)=a_{11}x_1^2+a_{12}x_1x_2+\cdots+a_{1n}x_1x_n \\ +a_{21}x_2x_1+a_{22}x_2^2+\cdots+a_{2n}x_2x_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ +a_{n1}x_nx_1+a_{n2}x_nx_2+\cdots+a_{nn}x_n^2.$$

一对相似的项，例如 $a_{12}x_1x_2$ 和 $a_{21}x_2x_1$，可以写成具同一系数的，使得它们每一个都是相应的变数乘积的系数的一半. 这样一来，每一个二次型自然地对应着它的系数组成的矩阵，它是对称矩阵.

二次型宜于用下面的矩阵写法表出. 记 X 为变数 $x_1, x_2, \cdots, x_n$ 所组成的列，$\overline{X}$ 为行 $(x_1, x_2, \cdots, x_n)$，即 X 的转置矩阵. 那么

$$f(x_1, x_2, \cdots, x_n)=x_1(a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n) \\ +x_2(a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n)+\cdots \\ +x_n(a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n)$$

$$=(x_1, x_2, \cdots, x_n)\begin{pmatrix} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n \\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n \end{pmatrix}$$

$$=(x_1, x_2, \cdots, x_n)\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=\overline{X}AX.$$

在很多数学部门及其应用中都会遇到二次型.

在数论和结晶学中所讨论的二次型，假定了变数 $x_1, x_2, \cdots, x_n$ 只取整数值. 在解析几何中二次型出现在 2 阶曲线(或曲面)的方程中. 在力学和物理中二次型作为用广义速度的分量表示系统的动能的表达式等等而出现. 此外，二次型的研究在分析中多变数函数的研究里，对于解决搞清楚已知函数在已知点的邻域中与逼近它的线性函数相差多少这类问题时是必要的. 研究函数的极大值及极小值就是这类问题的一个例子.

例如，现在来研究二个变数的函数 $w=f(x, y)$ 的极大值及极

小值的问题．假定函数 $w=f(x, y)$ 有直到3级的连续偏导数．(x_0, y_0) 点是函数 w 的极大值或极小值的必要条件是它的一级偏导数在 (x_0, y_0) 点等于零．假定这个条件成立，给了变数 x 和 y 的微小增量 h 和 k，来研究函数 w 相应的增量 $\Delta w=f(x_0+h, y_0+k)-f(x_0, y_0)$．根据泰勒公式，这个增量除了差一个高阶无穷小外，等于二次型 $\frac{1}{2}(rh^2+2shk+tk^2)$，其中 r, s, t 分别是二级偏导数 $\frac{\partial^2 w}{\partial x^2}$, $\frac{\partial^2 w}{\partial x \partial y}$, $\frac{\partial^2 w}{\partial y^2}$ 在点 (x_0, y_0) 的值．如果这个二次型对于 h 和 k 的一切值都是正的（$h=k=0$ 除外），那么函数 w 在 (x_0, y_0) 点有极小值；如果是负的，那么应在 (x_0, y_0) 点有极大值．最后，如果这个二次型既取正值又取负值，那么 w 在 (x_0, y_0) 点既无极大值也无极小值．同样可以研究更多个变数的函数．

二次型的研究，基本上在于研究在这一组或那一组线性变换下，二次型的等价问题．两个二次型称为是等价的，如果其中之一可以用已给线性变换集合中的一个变换变到另外的一个．与等价问题紧密相关联的是二次型的简化，即将它化为可能最简单的形状的问题．

在与二次型有关的各种问题里，也考虑各种不同的被容许线性变换的集合．

在分析问题里，要研究变数的任意非奇异线性变换；在解析几何中，最有兴趣的是正交变换，即对应着从一个笛卡儿坐标系到另一个笛卡儿坐标系的变换；最后，在数论和结晶学中，要研究带整系数的、具有行列式等于1的线性变换．

我们来研究两个这类的问题：将二次型用任意非奇异线性变换化为最简单形状的问题，以及用正交变换将它化为最简单形状的问题．首先我们要解释清楚，二次型的矩阵在变数的线性变换下如何改变．

设 $f(x_1, x_2, \cdots, x_n)=\overline{X}AX$，其中 A 是二次型的系数所组成的对称矩阵，X 是变数排成的列．

我们作一个变数的线性变换，将它简写作 $X=CX'$，这里 C

表示这个变换的系数矩阵，X' 表示新变数排成的列．那么 $\bar{X}=\bar{X}'\bar{C}$，因此 $\bar{X}AX=\bar{X}'(\bar{C}AC)X'$．所以，变换后的二次型的矩阵是 $\bar{C}AC$.

容易验证，矩阵 $\bar{C}AC$ 自然是对称的．因此，将二次型化为最简单的形状这个问题与将对称矩阵左方及右方乘以互为转置的非奇异矩阵而化为最简单的形状这个问题等价．

将二次型用逐步配方法化为标准形　我们要证明：任意（实的）二次型可用实的非奇异线性变换化为新变数的平方和而带有某些系数．

为了证明这个结果，我们先证明，如果这个二次型不恒等于零，那么可以作一个变数的非奇异线性变换使得第一个变数的平方的系数不等于零．

实际上，设

$$\begin{aligned}f(x_1, x_2, \cdots, x_n)=a_{11}x_1^2+a_{12}x_1x_2+\cdots+a_{1n}x_1x_n\\+a_{21}x_2x_1+a_{22}x_2^2+\cdots+a_{2n}x_2x_n\\\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\+a_{n1}x_nx_1+a_{n2}x_nx_2+\cdots+a_{nn}x_n^2.\end{aligned}$$

如果 $a_{11}\neq 0$，那么任何变换都不需要．如果 $a_{11}=0$ 而有一个对角线上的系数 $a_{kk}\neq 0$，那么令 $x_1=x'_k$，$x_k=x'_1$，并让其余原来的变数与相应的新变数相等．利用这个非奇异变换就能达到目的．最后，如果对角线的所有系数都等于零，那么一定有一个非对角线上的系数，例如 a_{12}，不等于零．

作非奇异线性变换

$$x_1=x'_1$$
$$x_2=x'_1+x'_2$$

而令其余原来的变数与相应的新变数相等，我们就达到目的．

因此，不失去普遍性我们可以假定 $a_{11}\neq 0$.

现在在二次型中分出来一个线性函数的平方，使得所有含 x_1 的项都出现在这个平方里．

这很容易做到．实际上，

$$f(x_1, x_2, \cdots, x_n) = a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n$$
$$+ a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n +$$
$$\cdots\cdots$$
$$+ a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2$$
$$= a_{11}\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n\right)^2$$
$$- a_{11}\left(\frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n\right)^2 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n +$$
$$\cdots\cdots$$
$$+ a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2.$$

将第二项的括号展开，然后归纳类似项，就得到

$$f(x_1, x_2, \cdots, x_n) = a_{11}\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n\right)^2 + f_1(x_2, \cdots, x_n),$$

其中 f_1 是 $n-1$ 个变数的二次型.

线性变换

$$\begin{aligned} x_1 + \frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n &= x_1' \\ x_2 &= x_2' \\ &\cdots\cdots \\ x_n &= x_n' \end{aligned}$$

显然是非奇异的. 作了这个变换之后，二次型就化为

$$a_{11}x_1'^2 + f_1(x_2', \cdots, x_n').$$

类似地把这个步骤进行下去，我们就将二次型化为所要求的“标准形”

$$\alpha_1 z_1^2 + \alpha_2 z_2^2 + \cdots + \alpha_n z_n^2,$$

这里 $z_1, z_2, \cdots, z_n$ 是最后引进的新变数.

二次型的惰性律 将二次型化为标准形时，在选取实现这个简化的变数变换时，有着很大的自由性. 这个自由性至少可以从这一点看出来，就是进行上面所叙述的逐步配方法之前，还可能将变数作任意的非奇异变换.

然而，纵然有这个自由性，不管所进行的变换是什么，最后几乎将二次型化为同一个标准二次型. 这就是说，新变数的平方项中带正系数的个数、带负系数的个数以及以零为系数的个数，是相同的. 这个定理称为二次型的惰性律. 我们不去证明它.

二次型的惰性律解决了实二次型对于一切非奇异变换的等价问题. 即，两个二次型是等价的，当且仅当将它们化为标准形后，所得到的标准形中带正系数的平方项的个数、带负系数的平方项的个数和以零为系数的平方项的个数都相等.

化为标准形后变成全是带正系数的新变数的平方和的二次型，在应用上特别有用. 这种二次型称为定正二次型.

定正二次型可以由下述性质来刻划，即对于将变数用不全为零的实数值代入二次型后，二次型的值永远是正的.

用正交变换将二次型化为标准形 将二次型化为标准形的所有可能的方法中，正交变换，即由变数的具正交矩阵的线性变换所实现的变换，特别有用. 例如在解析几何里将一个一般二次曲线或曲面的方程化简的问题中，正是这种变换有用处.

为了确信这种变换的可能性，最好将变数 $x_1, x_2, \cdots, x_n$ 看作是欧氏空间中一个变向量对于某组标准正交基的坐标，而将二次型看作是欧氏空间中向量的函数. 这时变数的正交变换就可以解释作从一组标准正交基到另一组标准正交基的转化.

将二次型

$$f(x_1, x_2, \cdots, x_n)=a_{11}x_1^2+\cdots+a_{1n}x_1x_n+\cdots\cdots+a_{n1}x_nx_1+\cdots+a_{nn}x_n^2$$

与线性变换 $\boldsymbol{A}$ 相联系，其中 $\boldsymbol{A}$ 在选定的基底中的矩阵是 $A=\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$，这时二次型本身就可以看作是数量积 $\boldsymbol{AX}\cdot\boldsymbol{X}$（其中 $\boldsymbol{X}$ 是坐标为 $x_1, x_2, \cdots, x_n$ 的向量），而系数 a_{ij} 就可以看作是数量积 $\boldsymbol{Ae}_i\cdot\boldsymbol{e}_j$，这里 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 是选定的那组标准正交基.

不难看出，由矩阵 $\boldsymbol{A}$ 的对称性推出对于任意向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$，等式

$$\boldsymbol{AX}\cdot\boldsymbol{Y}=\boldsymbol{X}\cdot\boldsymbol{AY}$$

成立.

首先证明，变换 $\boldsymbol{A}$ 至少有一个实特征根和对应它的一个特征向量.

为了这个目的，我们来研究二次型 $\boldsymbol{AX}\cdot\boldsymbol{X}$ 的值，假定 $\boldsymbol{X}$ 跑遍单位球，即跑遍全体单位向量的集合. 在这些条件下，二次型 $\boldsymbol{AX}\cdot\boldsymbol{X}$ 有极大值. 我们来证明，这个极大值 λ_1 就是变换 $\boldsymbol{A}$ 的特征值，而使 $\boldsymbol{AX}\cdot\boldsymbol{X}$ 达到这个极大值的向量 $\boldsymbol{X}_0$ 就是相应的特征向量，即 $\boldsymbol{AX}_0=\lambda_1\boldsymbol{X}_0$.

我们用间接方法来证明这个断言. 先来证明向量 $\boldsymbol{AX}_0$ 与一切与 $\boldsymbol{X}_0$ 正交的向量正交.

我们注意到，对于任意向量 $\boldsymbol{Z}$，不等式 $\boldsymbol{AZ}\cdot\boldsymbol{Z}\leqslant\lambda_1|\boldsymbol{Z}|^2$ 成立. 这是显然的，因为 $\boldsymbol{X}=\dfrac{\boldsymbol{Z}}{|\boldsymbol{Z}|}$ 是单位向量，而 λ_1 是 $\boldsymbol{AX}\cdot\boldsymbol{X}$ 在单位球上的极大值. 考察 $\boldsymbol{Z}=\boldsymbol{X}_0+\varepsilon\boldsymbol{Y}$，这里 ε 是一个实数，$\boldsymbol{Y}$ 是任意一个与 $\boldsymbol{X}_0$ 正交的向量，那么

$$\begin{aligned}\boldsymbol{AZ}\cdot\boldsymbol{Z}&=(\boldsymbol{AX}_0+\varepsilon\boldsymbol{AY})\cdot(\boldsymbol{X}_0+\varepsilon\boldsymbol{Y})\\&=\boldsymbol{AX}_0\cdot\boldsymbol{X}_0+2\varepsilon\boldsymbol{AX}_0\cdot\boldsymbol{Y}+\varepsilon^2\boldsymbol{AY}\cdot\boldsymbol{Y}\\&=\lambda_1+2\varepsilon\boldsymbol{AX}_0\cdot\boldsymbol{Y}+\varepsilon^2\boldsymbol{AY}\cdot\boldsymbol{Y}.\end{aligned}$$

此外，

$$\begin{aligned}|\boldsymbol{Z}|^2&=(\boldsymbol{X}_0+\varepsilon\boldsymbol{Y})\cdot(\boldsymbol{X}_0+\varepsilon\boldsymbol{Y})\\&=|\boldsymbol{X}_0|^2+\varepsilon^2|\boldsymbol{Y}|^2=1+\varepsilon^2|\boldsymbol{Y}|^2,\end{aligned}$$

因为

$$\boldsymbol{X}_0\cdot\boldsymbol{Y}=0,\ |\boldsymbol{X}_0|^2=1.$$

因此

$$\lambda_1+2\varepsilon\boldsymbol{AX}_0\cdot\boldsymbol{Y}+\varepsilon^2\boldsymbol{AY}\cdot\boldsymbol{Y}\leqslant\lambda_1+\varepsilon^2\lambda_1|\boldsymbol{Y}|^2,$$

除以 ε^2 后就得到

$$\frac{2}{\varepsilon}\boldsymbol{AX}_0\cdot\boldsymbol{Y}\leqslant\lambda_1|\boldsymbol{Y}|^2-\boldsymbol{AY}\cdot\boldsymbol{Y}. \tag{14}$$

最后，这个不等式对于任何实数 ε 都应该成立，而不管 ε 的绝对值多么小.

如果 $\boldsymbol{AX}_0\cdot\boldsymbol{Y}>0$，不等式(14)对于相当小的正数 ε 就不能成立；如果 $\boldsymbol{AX}_0\cdot\boldsymbol{Y}<0$，那么它对于绝对值相当小的负数 ε 就不能成立．因此，不等式(14)只有在 $\boldsymbol{AX}_0\cdot\boldsymbol{Y}=0$ 时才成立．这样一来，$\boldsymbol{AX}_0\cdot\boldsymbol{Y}=0$，即 $\boldsymbol{AX}_0$ 的确与 $\boldsymbol{X}_0$ 正交的向量正交．因此 $\boldsymbol{AX}_0$ 和 $\boldsymbol{X}_0$ 共线，即 $\boldsymbol{AX}_0=\lambda'\boldsymbol{X}_0$，其中 λ' 是一个实数．$\lambda'=\lambda_1$ 是容易验证的，即

$$\lambda_1=\boldsymbol{AX}_0\cdot\boldsymbol{X}_0=\lambda'\boldsymbol{X}_0\cdot\boldsymbol{X}_0=\lambda'.$$

现在容易证明，每一个二次型的确都可以用正交变换化为标准形．

设 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 是空间原来的一组标准正交基，$\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_n$ 是一组新的标准正交基，而第一个向量 $\boldsymbol{f}_1$ 等于变换 $\boldsymbol{A}$ 的特征向量 $\boldsymbol{X}_0$．设 $x_1, x_2, \cdots, x_n$ 是向量 $\boldsymbol{X}$ 在原来基中的坐标，而 $x_1', x_2', \cdots, x_n'$ 是它在新基中的坐标，那么

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=P\begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix},$$

其中 P 是正交矩阵．

对二次型行使转变到新变数的变换．对于新变数而言，二次型的系数 $a_{ij}'=\boldsymbol{Af}_i\cdot\boldsymbol{f}_j$．因此

$$a_{11}'=\boldsymbol{Af}_1\cdot\boldsymbol{f}_1=\lambda_1\boldsymbol{f}_1\cdot\boldsymbol{f}_1=\lambda_1,$$

$$a_{1j}'=a_{j1}'=\boldsymbol{Af}_1\cdot\boldsymbol{f}_j=\lambda_1\boldsymbol{f}_1\cdot\boldsymbol{f}_j=0,\ \text{当}\ j\neq 1\ \text{时},$$

即二次型的形状是

$$\lambda_1 x_1'^2+\varphi(x_2', \cdots, x_n').$$

因此，用了一个正交变换，我们在二次型中分出了一个新变数的平方．

将同样步骤行使到新二次型 $\varphi(x_2', \cdots, x_n')$ 上，然后再继续下去，最后我们就达到用一串正交变换将二次型化成标准形．显然一串正交变换等价于一个正交变换．定理得证．

§6. 矩阵函数和它的一些应用

矩阵函数 线性代数对其他数学部门的应用是很多的，也是各式各样的．说近代数学及理论物理的一大部分在这种或那种形式上采用了线性代数中的一些想法或结果，特别是矩阵计算，是并不夸张的．

我们来考虑将矩阵应用于常微分方程论的一个途径，这里矩阵函数起着主要的作用．

首先定义方阵 A 的方幂．令 $A^0=E$，$A^1=A$，$A^2=AA$，$A^3=A^2A$，$A^4=A^3A$ 等等．利用结合律容易证明，对任意自然数 m 和 n，有 $A^m\cdot A^n=A^{m+n}$．这使我们能够很自然地定义（一个变数的）矩阵多项式的值，即如果 $\varphi(x)=a_0x^n+a_1x^{n-1}+\cdots+a_n$，那么定义 $\varphi(A)=a_0A^n+a_1A^{n+1}+\cdots+a_nE$．这样，最简单的矩阵变数的函数——多项式函数就定义了．

利用极限的过程，我们很容易推广矩阵函数的概念到比一个变数的多项式更广泛的函数类上．我们不准备涉及这个问题最一般的情形，而只来研究解析函数．

首先我们引进矩阵序列的极限这个概念．矩阵序列

$$A_1=\begin{pmatrix} a_{11}^{(1)} & \cdots & a_{1n}^{(1)} \\ \cdots & \cdots & \cdots \\ a_{n1}^{(1)} & \cdots & a_{nn}^{(1)} \end{pmatrix},\quad A_2=\begin{pmatrix} a_{11}^{(2)} & \cdots & a_{1n}^{(2)} \\ \cdots & \cdots & \cdots \\ a_{n}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix},\ \cdots$$

称为收敛于矩阵 $A=\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$（或以 A 为极限），如果对一切 i，j 而言，有 $\lim\limits_{k\to\infty} a_{ij}^{(k)}=a_{ij}$．更进一步，级数 $A_1+A_2+\cdots+A_k+\cdots$ 的和定义为它的部分和的极限 $\lim\limits_{k\to\infty}(A_1+A_2+\cdots+A_k)$，如果这个极限存在的话．

设 $f(z)$ 是在 $z=0$ 的一个邻域内正则的解析函数．那么我们知道，$f(z)$ 可以展成幂级数

$$f(z)=a_0+a_1z+a_2z^2+\cdots+a_kz^k+\cdots.$$

对于任意方阵 A，自然可以令

$$f(A)=a_0E+a_1A+a_2A^2+\cdots+a_kA^k+\cdots.$$

有这样的结果，即如果矩阵 A 的特征值都落在幂级数 $a_0+a_1z+a_2z^2+\cdots+a_kz^k+\cdots$ 的收敛圆内，那么上面这个矩阵级数就收敛。

在应用中，矩阵的初等函数特别有用。

例如，几何级数 $E+A+A^2+\cdots+A^k+\cdots$ 对于特征值的模小于 1 的矩阵是收敛的，而这个级数的和是 $(E-A)^{-1}$，这与公式

$$1+x+x^2+\cdots+x^k+\cdots=\frac{1}{1-x}$$

完全相当。

把 $(E-A)^{-1}$ 表成无限级数这个事实，给出求系数矩阵接近于单位矩阵的线性方程组近似解的有效方法。

实际上，将这样一个方程组写作

$$(E-A)X=B,$$

我们得出

$$X=(E-A)^{-1}B=B+AB+A^2B+\cdots. \tag{15}$$

如果级数(15)收敛得很快的话，这给出解方程组的一个很方便的公式。

研究二项式级数

$$(E+A)^m=E+\frac{m}{1}A+\frac{m(m-1)}{2!}A^2+\cdots$$

是有用处的，这个级数中的指数 m 不只是自然数，而且也可以是分数或负数(如果 A 的特征值的模小于 1)。

对于应用特别重要的是矩阵的指数函数

$$e^A=E+A+\frac{A^2}{2!}+\frac{A^3}{3!}+\cdots.$$

用指数函数定义的级数，对于任何矩阵 A 都收敛。矩阵指数函数具有与通常指数函数相类似的性质。例如，如果 A 和 B 相乘是交换的，即 $AB=BA$，那么 $e^{A+B}=e^A\cdot e^B$。然而当 A 和 B 不交换

时，这个公式就不再成立了.

对线性常微分方程组理论的应用 在常微分方程组的理论中，考虑元素是一个独立变数的函数的矩阵

$$U(t)=\begin{pmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ \cdots & \cdots & \cdots \\ a_{m1}(t) & \cdots & a_{mn}(t) \end{pmatrix}$$

是有益处的.

对于这种矩阵可以自然地定义对于变数 t 的导数，即

$$\frac{dU(t)}{dt}=\begin{pmatrix} a'_{11}(t) & \cdots & a'_{1n}(t) \\ \cdots & \cdots & \cdots \\ a'_{m1}(t) & \cdots & a'_{mn}(t) \end{pmatrix}.$$

不难验证，有些基本微分公式对于矩阵成立. 例如

$$\frac{d(U+V)}{dt}=\frac{dU}{dt}+\frac{dV}{dt},$$

$$\frac{d(cU)}{dt}=c\frac{dU}{dt},$$

$$\frac{d(UV)}{dt}=\frac{dU}{dt}V+U\frac{dV}{dt}.$$

（乘法必须严格地按照公式中给出的次序进行！）

齐次线性常微分方程组

$$\frac{dy_1}{dt}=a_{11}(t)y_1+a_{12}(t)y_2+\cdots+a_{1n}(t)y_n,$$

$$\frac{dy_2}{dt}=a_{21}(t)y_1+a_{22}(t)y_2+\cdots+a_{2n}(t)y_n,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{dy_n}{dt}=a_{n1}(t)y_1+a_{n2}(t)y_2+\cdots+a_{nn}(t)y_n$$

用矩阵记法可以写成形状：

$$\frac{dY}{dt}=A(t)Y,$$

即写成类似于一个线性齐次微分方程的形状，其中

$$Y=\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},\quad A(t)=\begin{pmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ \cdots & \cdots & \cdots \\ a_{n1}(t) & \cdots & a_{nn}(t) \end{pmatrix}.$$

如果方程组的系数是常数，即矩阵 A 是常数矩阵，那么上面方程组的解可以像方程 $y'=ay$ 的解那样得到．即，这时 $Y=e^{At}C$，其中 C 是由常数组成的列．

这种形状的解研究起来很方便．这是因为对于任何解析函数 $f(z)$，等式

$$f(B^{-1}LB)=B^{-1}f(L)B$$

都成立．

因为任何矩阵都可以化为约当标准形（见§4），所以任意矩阵的函数的计算都可以化为标准矩阵的函数的计算，而这是容易办到的．因此，如果 $A=B^{-1}LB$，其中 L 是标准矩阵，那么

$$Y=e^{At}C=B^{-1}e^{Lt}BC=B^{-1}e^{Lt}C',$$

其中 $C'=BC$ 是任意常数组成的列．

由这个公式不难获得未知列 Y 的所有元素的明确表达式．

苏联学者拉波-达尼列夫斯基成功地发展了矩阵函数的理论，第一个将它应用于带变系数的方程组的研究．他的成果属于近五十年来数学的卓越成就之列．

文　　献

Ф. Р. 甘特马赫尔，矩阵论（上、下卷），高等教育出版社，1955. 本书包含线性代数应用方面的大量材料．

И. М. 盖尔冯德，一次代数学，商务印书馆，1953.

包括线性代数的几何叙述，从而进入泛函分析．

А. Г. 库罗什，高等代数教程，商务印书馆，1954.

А. И. 马力茨夫，线代数基础，商务印书馆，1954.

Л. Я. 奥库涅夫，高等代数（上、下册），商务印书馆，1955.

В. И. 斯米尔诺夫，高等数学教程（第三卷），高等教育出版社，1954.

В. Н. 法捷也娃，线性代数计算法，科学出版社，1958.

阐述了线性代数基本问题的数值解法．

万哲先 译　刘绍学 校

第十七章 抽象空间

自从罗巴切夫斯基最先证明了非欧几里得几何的可能性和提出了几何学对于物质现实的关系的新的看法以来，几何学的对象、方法和应用大大地扩大了．现在数学家们在研究着各种的“空间”，不但研究欧几里得空间，同时还研究罗巴切夫斯基空间、射影空间、各种 n 维空间以至无穷维空间、黎曼空间、拓扑空间以及其它的空间．这种空间的数目无穷，而且其中每一种空间都有自己的性质、自己的“几何”．在物理学中使用着所谓相“空间”和位形“空间”；相对论要用到空间的曲率观念和抽象的几何理论的其它结论．

这些数学的抽象究竟是如何产生又从何而产生的呢？它们又有什么现实基础、现实意义和应用呢？它们与现实的关系是怎样的呢？在数学中是如何定义它们和讨论它们的呢？现代几何学的普遍观点在数学中有什么意义呢？

这些问题就是本章所要回答的．但是在本章中不预备叙述抽象的数学空间的理论本身，因为这样做需要非常多的篇幅和在专门的数学工具上化极大的功夫．我们的任务是阐明几何学的新观念的本质，即回答上面所提的问题．这都可以不用复杂的证明和公式而做到．

问题的历史回溯到欧几里得的“几何原本”，回溯到关于平行线的公理（或说公设）．

§1. 欧几里得公设的历史

欧几里得在他的“几何原本”中以所谓公设和公理的形式写出了几何学的一些基本的前提．其中包含着第五条公设（在“几何原

本"的另一些版本中是第十一条公理），这公设在现在普通是这样叙述的："通过不在已知直线上的一个点，不能引多于一条的直线，平行于已知直线。"我们回想一下，所谓一条直线平行于一条已知直线，是指这两条直线在一个平面上而且不相交，这时说的是无限的直线而不是它们的有限线段。

容易证明：通过不在已知直线 a 上的点 A，至少总可以引一条直线平行于已知直线。

实际上，我们从点 A 引垂直线 b 到直线 a 上，而且通过 A 引直线 c 垂直于 b(图 1)。得到的图形正好对称于直线 b，因为直线 b 在其两侧与直线 a 和 c 组成的角相等。因此，当沿着直线 b 把平面翻转过来时，我们就把直线 a 和 c 的两半交换了位置。由此可知，假如 a 和 c 在 AB 的某一侧相交，那末它们也应该在另一侧相交。在这种情况下，直线 a 和 c 将有两个公共点，然而这是不可能的，因为按照直线的基本性质，通过两个点只可以引一条直线(有两个公共点的直线必然重合)。

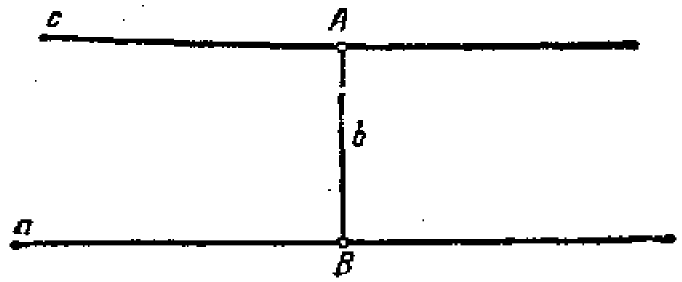


图 1

总之，从直线的基本性质和图形的移动(由于沿着直线 AB 翻转平面就是绕着这条直线来旋转半平面)推出，通过已知点至少总能引一条直线平行于已知直线。这个结论在补充了欧几里得公设以后，就可以断言这种平行线只有一条，而不能有别的平行线。

在几何学的其他公设(公理)中，这个公设占有某种独特的地位。欧几里得自己把这个公设写得非常复杂，即使在上面所引的普通的形式中，这公设也包含着一定的困难之处。这个困难就包括在平行线概念本身之中。这里谈到了整条直线。但是如何断定两条已知直线平行呢？为此必须把它们向两侧引到"无穷"，而且要断

定它们在无限延长的整个范围内没有一处是相交的．显然，这种想法有其困难之处．欧几里得自己就已经让平行线公设占有某种独特的地位，看来完全是有理由的了．在欧几里得的“几何原本”里，这个公设从第 29 个命题才开始应用，同时欧几里得是不用这个公设而得出前 28 个命题的．由于这公设的复杂性，很自然地产生能否不用它的愿望，因此早在古代就已经出现了修改平行线的定义、修改公设的表述的企图，甚至更进一步把它作为定理而从几何学的其他公理和基本概念导出的企图．

这样，基于第五条公设的平行线理论从古代开始就成为许多数学家工作中的注释和研究的对象．在这一连串的研究中，首要的任务是完全摆脱第五条公设，把它作为定理而从几何学的其他基本命题导出．

从事这个工作的有很多几何学家：为欧几里得作注解的希腊的普洛克尔（公元五世纪），伊朗的纳西艾丁·屠西（十三世纪），英国的瓦里斯（1616—1703），意大利的萨克利（1667—1733），德国的哲学家兼数学家兰贝尔特（1728—1777），法国的勒让德（1752—1833）和其他许多人．在欧几里得的“几何原本”问世以后的两千多年长时期中，他们全都具有细心和几何的机智而企图证明第五条公设．

然而这些努力肯定地都得到相反的结果．每一次都显示出，这个或那个证明的作者实际上却运用了某个命题，这个命题可能是显然的，但是却完全不能以逻辑的必然性从几何学的其他前提推得．换句话说，每一次的情况都是把第五条公设换成别一个命题，从这个命题确实能推得这个公设，但是这个命题本身却有待于证明[1)]．

1）这种等价于第五条公设的命题可以提出很多．下面就是一些例子：1）平行于已知直线的直线，与已知直线有定常的距离（普洛克尔）；2）存在着相似（而不相等）的三角形，即角相等而边不相等的三角形（瓦里斯）；3）至少存在一个长方形，即四个角都是直角的四角形（萨克利）；4）垂直于锐角的一条边的直线，也与这锐角的另一条边相交（勒让德）；5）三角形各角之和等于两直角（勒让德）；6）存在着面积任意大的三角形（高斯）．这样的例子我们现在可以无限制地列举下去．

从另一方面深入到问题之中的是萨克利和兰贝尔特. 萨克利第一个企图用反证法来证明第五条公设, 即他在起初采用相反的断言, 然后从它展开推论, 期望引向矛盾. 在这些推导中达到了一些完全无法思议的结果, 他就以为问题解决了. 但是错了, 因为与直观表象的矛盾并不同时表明逻辑的矛盾. 要知道问题在于根据几何学的其他命题来确立欧几里得公设的逻辑证明, 而不是为了重新肯定它的直观的正确性. 这个公设就其本身说在直观上是十分令人相信的. 但是, 让我们再说一遍, 直观的信服和逻辑的必然性是两回事.

兰贝尔特是比萨克利和他的先驱者更为深入的思想家. 他沿着同样的道路前进, 但是既未引出逻辑的矛盾, 也未做出其他的错误; 他并没有声明他似乎证明了第五条公设. 但是在他以后, 到了十九世纪的开端, 勒让德重又"证明"了第五条公设. 他犯的是老错误, 他仍然是用其他有待于证明的断言代替了这个公设.

总之, 直到十九世纪开始时, 第五条公设的证明问题还是没有解决, 就象在欧几里得时代一样. 白费了许多人的精力, 问题未见改善. 这真是几何学的一个深奥的谜, 一个看来毫无疑问能被优秀的几何学家解决的问题, 却在两千年的长时期内未能得到解决.

平行线理论在十九世纪成为几何学的中心问题之一. 研究它的有很多几何学家: 高斯, 拉格朗日, 达朗贝尔, 勒让德, 瓦赫特, 史威卡特, 塔乌里努斯, 法尔卡什·伯依阿依和其他人.

然而公设的证明并未得到. 这是怎么回事呢? 是没有办法解决问题呢, 还是问题的提法就不对呢? 这问题已经开始出现在以思考深入而胜过他人的某些数学家面前. 著名的德国数学家高斯从1792年起就致力于这个问题, 而且问题的正确提法逐渐显现在他的眼前. 最后, 他决定丢弃第五条公设, 而且从1813年起发展了从相反的断言推导出来的一系列定理. 不久以后, 德国数学家史威卡特当他在哈尔科夫担任法学教授时也走上了同一条道路, 后来塔乌里努斯也是如此. 但是在他们中谁都没有最终地找到问题的答案. 高斯细心地隐藏了自己的研究, 史威卡特只是给高斯一封

私人信，只有塔乌里努斯刊行了以第五条公设的否定为基础的新几何学的初步．然而他自己也排斥了这种几何学的可能性．因此，他们之中谁都没有解决问题，而且关于这问题的整个提法的正确性的问题还是没有答案．在1826年2月23日喀山大学的青年教授罗巴切夫斯基才最先给出答案．他在物理数学系的会议上宣读了关于平行线的报告，这篇报告的内容在1829年刊登在喀山大学学报上．

§2. 罗巴切夫斯基的解答

1．罗巴切夫斯基在他的著作“新几何原本”(1835)里，以下列文句描述了他所给出的第五条公设的解答要点：

“大家知道，直到今天为止，几何学中的平行线理论还是不完全的．从欧几里得时代以来，两千年来的徒劳无益的努力，促使我怀疑在概念本身之中并未包括那样的真情实况，它是大家想要证明的，也是可以像别的物理规律一样单用实验(譬如天文观测)来检验的．最后，我肯定了我的推测的真实性，而且认为困难的问题完全解决了．我在1826年写出了关于这个问题的论证．”

我们来分析一下，罗巴切夫斯基在这一段话里谈的是些什么．这段话像焦点一样地集中了罗巴切夫斯基的新观点，在其中不仅给出了关于第五条公设的问题的解答，而且使几何学的全部注意力转移到新的方面，甚至还不单几何学是如此．

罗巴切夫斯基早在1815年就开始研究平行线理论，他最初也像其他几何学家一样，企图证明第五条公设．在1823年他已经清楚地意识到，所有的证明，“不论是如何地给出的，只可以认为是说明，而在数学证明的完整意义下是不应当获得尊重”的[1]．在那儿他已经看到，“在概念本身之中并未包含大家想要证明的真情实况”，换句话说，从几何学的基本的前提和概念并不能推导出第五

1) 这是罗巴切夫斯基于1823年在他的几何学教程里所写的．这本书并未在他生前刊行．“几何学”教程直到1910年才出第一版．

条公设．那么他怎么肯定这种推导的不可能性呢？

他肯定了的是，可以循着萨克利和兰贝尔特曾经走过前几步的途径，继续地走下去．他引用了与欧几里得公设相反的断言，即“通过不在已知直线上的点，可以引不止一条而至少是两条直线，平行于已知直线”作为假设．当我们约定采用这个断言作为公理，而且把它与几何学的其他命题连结起来以后，我们就可以由此展开极深入的推论．于是，假如这种断言与几何学的其他命题不相容，我们就会引向矛盾，因而第五条公设也就用反证法证明了：否定命题引向了矛盾．然而并未出现这种矛盾，我们就引出了罗巴切夫斯基曾经作过的两个结论．

第一个结论是：第五条公设不能证明．第二个结论是：在刚才所写的否定公理上可以展开一连串的推论——定理，这些定理并不包含矛盾．因而这些推论就形成了一个逻辑上可能的、无矛盾的理论，这个理论可以看做是新的非欧几里得的几何学．由于还不能找出这种几何的现实意义，罗巴切夫斯基慎重地把它叫做“虚拟的几何学”．但是他已经看到了它的逻辑的可能性．在说出了和坚持了这个巩固的信念时，罗巴切夫斯基表现出是真正伟大的天才，他毫不动摇地坚持了自己的信念，并不因为社会的舆论，也并不因为怕遭受误解和批评而隐藏这些信念．

总之，罗巴切夫斯基所得到的前两个结论是：断定了第五条公设的不可证明性和根据否定公理展开新几何学的可能性．假如不考虑它的结论与关于空间的直观看法的矛盾性，这新的几何学在逻辑上像欧几里得几何学一样是丰富而又完善的几何学．罗巴切夫斯基发展了这个新几何学，因此现在这种几何学就以他的名字来命名．这里包含着极为重要的一个普遍结果：**在逻辑上可能的并不只是一种几何学**．我们还将说明这个结论所包含的全部意义；特别说来，回想到本章开头所提出的关于抽象的数学空间的诸问题，在这个结论里已经包含了那些问题的不小一部分解答了．

我们再回到上面提到过的罗巴切夫斯基的话．他说几何的真理，像其他的物理规律一样，可以单用实验来检验．这首先是说，

所谓真理应该理解为抽象概念与现实的一致性. 这种一致性可以单用实验来确定. 因此,为了检验这些或那些结论的真实性,必须作试验性的研究,单靠逻辑的推断对于这个是不够的. 虽然欧几里得几何很精确地反映了空间的现实性质,但是要说即使作进一步的研究,也不能发现欧几里得几何作为关于现实空间的性质的学说只是近似地正确的,这是无法使人信服的. 于是几何学作为关于现实空间的学说(而不是作为逻辑系统),就需要根据新的实验数据加以修正和精确化了.

罗巴切夫斯基的这个天才的思想,在物理学的新的发展中——在相对论中,得到了完全的证实.

罗巴切夫斯基自己根据天文观测来作计算,以便检验欧几里得几何的精确性. 这些计算当时在可达到的精确性范围内证实了欧几里得几何的正确性. 现在情况变了,虽然应该立即指出,在应用到空间时罗巴切夫斯基几何并不显得更精确些,空间显得还有其他的更复杂的性质. 但是罗巴切夫斯基几何早就通过其他的关系获得了自己的根据和应用,这关系我们在以后将会详细地说到.

必须指出,罗巴切夫斯基完全没有简单地把他的几何看作建立在任意采用的假设上的逻辑系统. 他看到主要的问题不在于几何基础的逻辑分析,而在于它们与现实的关系的研究. 由于实验不能给出欧几里得公设真实性问题的绝对地精确的解,所以研究作为几何学的更基本的前提的逻辑可能性才有了意义. 这种数学的研究促成了对空间的现实性质作物理的研究时所应该遵循的途径,何况欧几里得几何还是罗巴切夫斯基几何的极限情形,因而这后者包括了更大的可能性. 从这个观点看来,欧几里得公设的限制等于是对于理论发展的禁令. 理论必须超出早为大家所知道的界限,才能找到和指出发掘新事实和新规律的途径. 对于数学与现实的联系的深入理解,可以从现实的多样性中分出那样一些逻辑的可能性,它有在自然知识中显得合用的最广的基础. 假如追随罗巴切夫斯基的几何学家们没有发展关于空间的可能性质的学说,那末现代物理学家就不能有那样一些数学的工具,使它能利用

这些工具来表述和发展相对论的命题了.

总之，我们可以来总结一下罗巴切夫斯基所给出的第五条公设问题的解答.

1°. 公设是不能证明的.

2°. 几何学的其他基本命题添上否定公理以后，可以展开一种与欧几里得几何不同的、逻辑上完整而富有内容的几何学.

3°. 这种或那种逻辑上可能的几何学的结论，在应用到现实空间时的正确性只有用实验来作检验. 逻辑上可能的几何学不应该当作任意的逻辑体系来研究，而应该作为促成发展物理理论的可能途径和方法的理论来研究.

这个解答与企图证明欧几里得公设的几何学家们所希望获得的解答完全不同. 这个解答与早已确立的观念是如此地背道而驰，以致于在数学家中间也不能被理解. 它对于他们说已经是太新、太激进了. 罗巴切夫斯基就像是砍断了平行线理论中的郭尔迪亚王之结*，而不是像其他几何学家所想做到的那样来解开它.

2. 几乎与罗巴切夫斯基同时，匈牙利的几何学家雅诺什·伯依阿依(1802—1860)也发现了第五条公设的不可证明性和非欧几里得几何的可能性，雅诺什·伯依阿依在他父亲法尔卡什·伯依阿依的1832年出版的几何论文中，以附录的形式发表了他的结论. 事先做父亲的曾把儿子的工作寄给高斯评阅，得到了高斯的称赞，高斯说他自己在前些时也获得了同样的结论. 然而高斯始终没有把这种结论放在他的论著里发表. 在一封信中他解释说，是因为恐怕不被人所理解.

在科学中经常发生这样的事，在科学中的一些成熟了的结论几乎同时地和独立地为不同的学者所获得. 牛顿和莱布尼茨同时

* 关于郭尔迪亚王之结的传说是：有预言说能解开这个结的就能统治小亚细亚，然而始终无人解开. 后来马其顿的阿历山大王听到这个预言，用剑把这个结砍成两断. 本句话的意思是说，罗巴切夫斯基用出人意料的办法解决了平行线理论中的难题.——译者注

发展了微积分法；达尔文的学说同时被渥莱斯所获得；与爱因斯坦同时树立相对论原理的还有庞加莱．还可以举出很多这样的例子．它们再一次地证明了，科学的发展必须通过在科学中成熟了的问题的解决，而不是通过那些偶然的发现和猜测．就这样同时发现非欧几里得几何的可能性的有好几个几何学家：罗巴切夫斯基、伯依阿依、高斯、史威卡特和塔乌里努斯．

然而，在科学中并非所有获得新成果的学者在确立这种新成果时都起着同样的作用，因而不能给他们以同等的功绩．这时还要考虑到时间的优先性、结论的明白性和深度、它们产生的顺序和根据．不论史威卡特还是塔乌里努斯，都没有确信几何学的合法性．而在现在这个情形里，这正是有决定意义的，尤其因为萨克利和兰贝尔特已经得到过新几何学的一些个别的结论．高斯虽然好像有了这种确信，但是并未如此地坚持这种确信，以致于肯冒险把它发表出来．还没有一个深思的和多才的数学家，能像罗巴切夫斯基所做的那样，勇敢地坚持了新的思想，以致于在数学中引起了革命．在这一点上，高斯是那个时代的德国知识分子的典型代表者，这些知识分子同时具有深入的理论知识和政治上的软弱性．这种精神表现在高斯的同时代人黑格尔的哲学之中．列宁曾经说过，黑格尔"天才地预见了，也只是预见了"自然和认识的辩证法，但是在预见了这个赫尔岑所谓"革命的代数学"以后，他自己却使它从属于唯心主义的、反动的哲学体系．

伯依阿依虽无不坚决的表现，但是他所给出的新的观念，并不像罗巴切夫斯基那样深入地展开．正是罗巴切夫斯基第一个公开地——1826年在口头上，1829年在文字上——提出了新的观念，而且继续在一系列的工作中发展了和宣传了这种观念，最后在1855年刊行了"泛几何学"．他在口述这个著作时，已经是暮年时代的瞎眼老人了，但是仍然保持着坚强的精神和确信自己的正确性．正因为如此，所以理应以他的名字来称呼新的几何学．

罗巴切夫斯基不仅发展了新几何学，而且还正确地提出了关

于几何学与现实的关系的问题．哲学家康德直到今天还在唯心主义者中享有很大的荣誉，就因为他认为空间不是物质存在的现实形式，而只是我们的观念的天赋形式，想象的先验的（即与经验无关的）形式．因此，照康德的说法，几何学也是先验的，即与经验无关的．罗巴切夫斯基驳斥了这种唯心主义的见解，他作为唯物主义者而站在康德的对立面，提出了几何学与现实的关系的唯物主义的解释．他说，真理只能用经验来检验．与康德相反，罗巴切夫斯基断言，“不应该信赖通过天赋的感觉而获得的初始的（即基本的）概念”．罗巴切夫斯基不仅坚持认为观念和概念都从经验产生（这是他以前的唯物主义者所说的），而且还指出，几何学与现实的关系也必须用经验来使它精确化．这已经包括了发展几何学的思想，包括了随着我们知识的发展无限制地逼近于绝对真理的思想．

因此，罗巴切夫斯基不仅是天才的几何学家，而且是唯物主义的哲学家．他还是俄国教育界的精力充沛的和多才多艺的活动家，他在喀山大学担任教授而且当了将近20年的校长．罗巴切夫斯基生活在坚持自己的思想、不管别人的怀疑甚至嘲笑的唯物主义、勇敢和刚毅之中，生活在作为学者、教育家和组织家的广泛活动之中，表现出当时俄国教育界的优秀代表的特色．当时正是俄国天才和社会意识暴风雨般地兴起和隆盛的时代，正是普希金、十二月党人、别林斯基的时代．俄国并不像欧洲文化的一个胆怯的和顺从的女学生而出现在国际舞台上，而是作为能给出自己的、新的、而且还是欧洲所不知道的那种东西的力量．就这样，罗巴切夫斯基在数学中解决了两千年来屹立在科学面前的问题，而且使科学向新的道路上发展．他的唯物主义和他的科学的勇气类似于拉其晓夫、彼斯捷尔和别林斯基的唯物主义和科学的勇气．罗巴切夫斯基不仅是新几何学的创立者，而且是学者、思想家和公民——伟大的人（在这个字的整个意义下）．在罗巴切夫斯基死后，几何学的发展当然远远超出了他所能想象到的范围，但是实在应该认为几何学以致一般数学的新时代是从他开始的．

§3. 罗巴切夫斯基几何

1. 总之，罗巴切夫斯基采用了与第五条公设相反的断言作为基础：在已知平面上通过一个点至少可以引两条直线不与已知直线相交．他由此导出一系列深入的推论，它们就组成新的几何学．因此，这种几何学是作为某种想象的理论，作为从上面所作的假设与欧几里得几何（罗巴切夫斯基说它是"通常的"几何）的其他[1]基本前提出发而用逻辑方法证明的定理的总体而建立的．

罗巴切夫斯基在他的推导里得到了与"通常的"初等几何相类似的全部结果，即达到了非欧几里得的三角学和三角形的解法，达到了面积和体积的计算．我们在这里不能追踪罗巴切夫斯基的推导的这个线索，并非由于它们过分复杂，而首先是因为没有足够的篇幅．要知道"通常的"几何的中学课本是相当厚的，而罗巴切夫斯基的推导当然不会比这些"通常的"推导更简单些更短些，所以我们在这里只能提出罗巴切夫斯基几何的一些突出的结果．有兴趣更深入研究非欧几里得几何的读者，可以去看专门的著作．然而我们将要比较深入地说明非欧几里得几何的简单的现实意义．

我们从平行线理论开始．设给了直线 a 和直线外的点 A，从 A 引垂直线 AB 到直线 a 上．按照基本的假设，至少存在两条直线，通过点 A 而且不与直线 a 相交．于是处在这两条直线的交角里的每一条直线也都不与 a 相交．在图 2 上直线 b 和 b' 在延长后与 a 相交是违反罗巴切夫斯基的假设的．但是这并不足怪．要知道罗巴切夫斯基所论证的不是我们在普通平面上所作的图，他是从他的假设来作逻辑的推导的，这个假设是与我们惯常在图上看到的情形

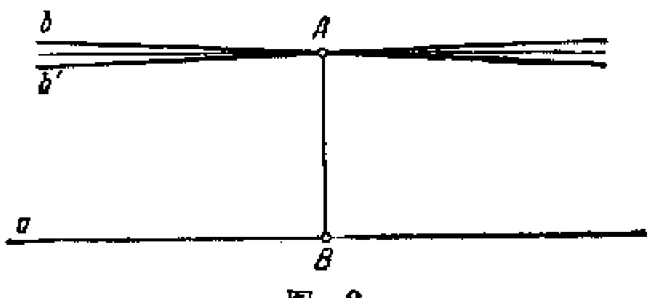


图 2

1) 几何学的这些所谓"其余的"命题，以后（在§5里）还要精确地表述．

相对立的. 图在这里只起着辅助的作用, 在图上并没有精确地画出非欧几里得几何的事实, 因为我们在图上所画的是普通平面上的普通的直线, 这在图的精确范围内无疑是属于欧几里得几何的.

在逻辑的可能性和直观的表现之间的这个矛盾是理解罗巴切夫斯基几何的主要困难. 但是假如谈的是作为逻辑理论的几何学, 则应该考虑的是论证的逻辑严密性, 而不是与惯常的图象是否协调.

2. 我们重新回到直线 a 和点 A. 从 A 引不与 a 相交的半直线 x(例如垂直于 AB 的半直线), 而且让它绕着点 A 旋转, 使得 AB 和 x 的夹角 φ 变小, 但是并不使这条半直线与 a 相交. 于是半直线 x 就趋向一个极限位置, 对应于角 φ 的最小的值, 这个极限的半直线 c 也不与 a 相交.

实际上, 如果它与直线 a 相交于某个点 X (图 3), 则我们可以在右边取点 X' 而得到半直线 AX', 它与 a 相交, 但是与 AB 组成较大的角. 这是不可能的, 因为按照半直线 c 的构成, 与 AB 组成较大的角的半直线 x 都不应该与直线 a 相交.

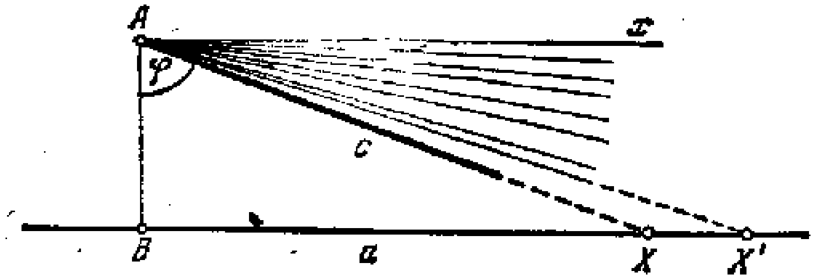


图 3

因此, 半直线 c 不与 a 相交, 并且在通过点 A 而且不与直线 a 相交的所有半直线中, 它是边界.

按照对称性, 显然在另一侧也可以引一条半直线 c', 它不与 a 相交而且是所有这种半直线的边界. 假如 c 和 c' 彼此是延长线, 那末它们就共同组成一条直线 $c+c'$. 于是这条直线就是通过已知点 A 而且平行于 a 的唯一的直线, 因为当它稍微旋转一下时, 或是 c 或是 c' 就要与 a 相交. 只要一假定平行线不是一条, 而是至少有两条, 则半直线 c 和 c' 就不是彼此延长的了.

总之，我们证明了罗巴切夫斯基几何的第一个定理．

从不在已知直线 a 上的点 A，可以引两条半直线 c 和 c'，使得它们不与直线 a 相交，但是处在它们的夹角之中的任何半直线都与直线 a 相交．

假如把半直线 c 和 c' 延长，则我们就得到两条直线(图 4)，它们都不与 a 相交而且具有这样的性质：在这两条直线的夹角 α 里，通过点 A 的每一条直线都不与直线 a 相交，而在角 β 里通过的每一条直线则都与直线 a 相交．罗巴切夫斯基把这种直线 c 和 c' 叫做直线 a 的平行线，直线 c 叫做右平行线，直线 c' 叫做左平行线．角 β 的一半，罗巴切夫斯基叫做平行角；因为角 β 小于两直角，平行角小于直角．

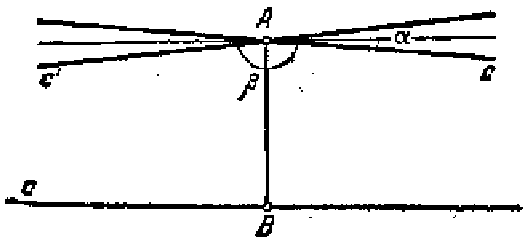


图 4

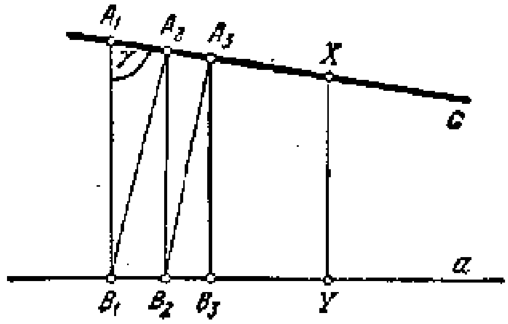


图 5

3．现在我们来察看一下，从直线 c 上的点 X 到直线 a 的距离在 X 沿着 c 移动时(图 5)如何改变．在欧几里得几何里，平行直线之间的距离是常数．而在这里我们要肯定的是，当点 X 向右移动时，它到 a 的距离(即垂直线段 XY 的长度)变小．

我们从点 A_1 引垂直线 A_1B_1 到直线 a 上．我们再从点 B_1 引垂直线 B_1A_2 到直线 c 上(点 A_2 在点 A_1 的右侧，因为角 γ 是锐角)．最后，我们从点 A_2 引垂直线 A_2B_2 到直线 a 上．我们来证明 A_2B_2 小于 A_1B_1．

垂直线短于倾斜线的定理在罗巴切夫斯基几何里也成立，因为这定理的证明(可以在任何一本中学几何教本里找到)并不运用平行线的概念和与它有关的推导．只要垂直线短于倾斜线，则作为直线 c 的垂直线的 B_1A_2 就短于 A_1B_1，同理，作为直线 a 的垂

直线的 A_2B_2 短于 B_1A_2. 因此 A_2B_2 短于 A_1B_1.

其次，从点 B_2 引垂直线 B_2A_3 到直线 c 上而且重复同样的论证，可以肯定 A_3B_3 短于 A_2B_2. 继续这个作图法，我们得到越来越短的垂直线序列，即点 A_1, A_2, … 到直线 a 的距离是变小的. 然后，我们补充了简单的论证以后，就可以证明，一般地说，如果在 c 上点 X'' 处在 X' 之后，则垂直线 $X''Y''$ 短于 $X'Y'$. 我们不再停留在这一点上. 我们希望以上的论证已经足以说明事情的实质，而我们的任务并不在于严密的证明.

但是值得指出，可以证明，距离 XY 当点 X 沿着直线 c 向右移动时不仅变小，而且它在点 X 趋向无限远处时还要趋向于零. 这就是说，平行直线 a 和 c 渐近地逼近！ 同时可以证明，它们之间的距离在相反的方向不仅增大，而且趋向无穷.

在欧几里得几何里，平行于已知直线的直线与它有定常的距离. 而在罗巴切夫斯基几何里，一般地不存在这样一对直线，在那儿两条直线总要分离到无穷，不是在一方面，就是在两方面. 至于与已知直线有定常距离的线，那总不是直线，而是一种曲线，叫做等距线.

罗巴切夫斯基的这些结论确实是骇人听闻的和完全不与惯常的直观表象一致的. 但是正像我们已经说过的，这种不协调并不能作为一种论据，来否定罗巴切夫斯基几何是从所采用的前提作

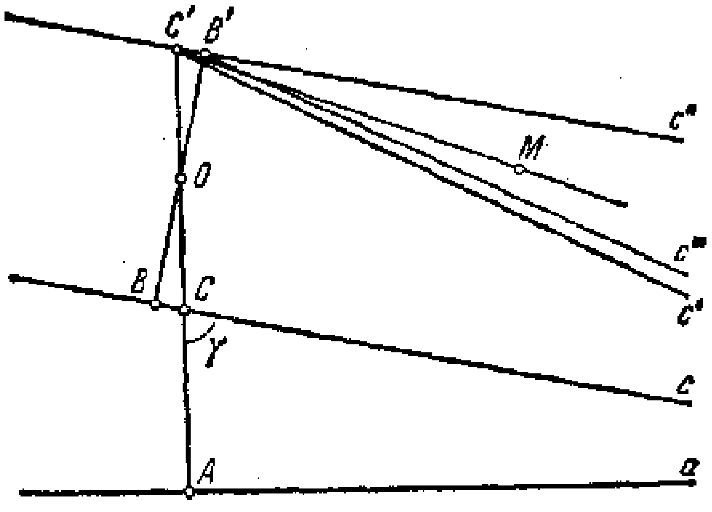


图 6

逻辑推演的抽象理论.

4. 现在我们再来察看平行角，即由平行于已知直线 a 的直线 c 与垂直线 CA 组成的角 γ(图 6). 我们要证明，当点 C 离开直线 a 越远时，这个角也就越小. 为此我们先证明下列事实：如果两条直线 b 和 b' 与截线 BB' 组成相等的角 α 和 α'，则这两条直线有公共的垂直线(图 7).

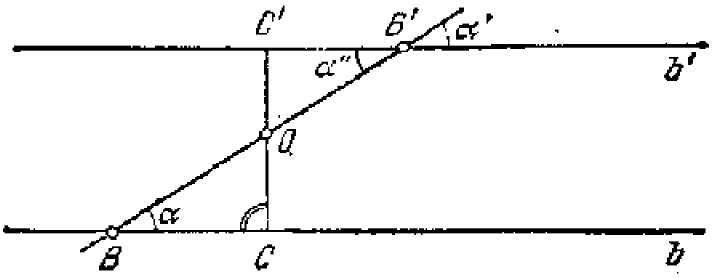


图 7

为了证明，我们通过线段 BB' 的中点 O 引垂直于直线 b 的直线 CC'，我们得到两个三角形 OBC 和 $OB'C'$. 按照作图，这两个三角形的边 OB 和 OB' 相等. 在公共顶点处的角是对顶角，也相等. 角 α'' 等于角 α'，因为它们也是对顶角. 按照已知条件，角 α 等于角 α'. 因此角 α 也等于角 α''. 这样一来，在我们的三角形 OBC 和 $OB'C'$ 中，边 OB 和 OB' 以及与它们邻接的角都相等. 于是根据熟知的定理，两个三角形相等，特别地，它们在点 C 和 C' 处的角也相等. 但是角 C 是直角，因为按照作图，直线 CC' 垂直于 b. 因此角 C' 也是直角，即 CC' 也垂直于直线 b'. 这样一来，线段 CC' 就是两条直线 b 和 b' 的公共的垂直线. 公共垂直线的存在也就证明了.

现在我们来证明，平行角随着点到直线的距离的变大而变小. 这就是说，如果点 C' 到直线 a 的距离远过于 C(像在图 6 上所画的那样)，则从 C' 所引的平行线 c' 与垂直线 $C'A$ 组成的角就要小于从 C 所画的平行线 c 的角.

为了证明，我们从 C' 引直线 c''，使它与 $C'A$ 组成的角等于平行线 c 与 CA 组成的角. 于是直线 c 和 c'' 与截线 CC' 组成相等的角，因而像刚才所证明的那样，它们有公共的垂直线 BB'. 于是

从这条垂直线的垂足 B' 可以引直线 c'''，使它平行于 c 而且与垂直线组成小于直角的角，因为我们知道，平行线与垂直线组成小于直角的角．现在我们在直线 c'' 和 c''' 的夹角里取任意点 M 而且引直线 $C'M$．它进入 c'' 和 c''' 的夹角而且以后再不能与 c''' 相交．因而它也不会与直线 c 相交．但是它与 AC' 组成的角已经比 c'' 与 AC' 组成的角为小，即小于角 γ．其次，更小的角是由平行线 c' 形成的，因为它是通过 C' 而且不与 a 相交的所有直线中的边界．因此，平行线 c' 与 $C'A$ 组成的角比 c 为小，这就表明，平行角在变得更远的点 C' 处变得更小，这就是所要证明的．

总之，我们证明了平行角随着点 C 离开直线 a 变远而变小．但是原来还可以更进一步地证明：如果点 C 趋向无穷远处，则这个角趋向于零．这就是说，当点到直线 a 的距离相当大时，平行线与这直线的垂直线组成任意小的角[1]．换句话说，如果在离直线 a 很远的地方，作与这直线的垂直线组成很小倾角的直线，则我们可以沿着这条"倾斜的"直线而永远不与直线 a 相交．罗巴切夫斯基几何的这个事实也造成令人惊奇的印象．但是以后还可以得出别的更惊奇的结果．

举例说，我们取由半直线 a 和 a' 组成的锐角 α，在离角 α 的顶点相当远的地方引 a 的垂直线 b，使得对应于所取的距离 OB 的平行角小于 α（图 8）．因为角 α 大于平行角，所以通过点 O 而且平行于直线 b 的直线 b' 与 a 组成较小的角．但是直线 b' 不与直线 b 相交，因此 a' 更不与 b 相交．这就证明了，在锐角一边上离顶点足够远的地方所引的垂直线不与另一边相交．

5．我们引出以上的全部推导有两个目的：第一（这是主要的），我们希望指出一些最简单的例子，来说明怎样从所采用的前提出发可以得出罗巴切夫斯基几何的定理．这可以作为最简单的

1）如果 h 是垂直线的长度，γ 是平行角，则罗巴切夫斯基证明了 $\mathrm{tg}\,\frac{\gamma}{2}=e^{-\frac{h}{k}}$，这里 k 是依赖于长度单位的常数，e 是熟知的自然对数的底数．明显地，当 $h\to\infty$ 时，$e^{-\frac{h}{k}}$（因而 γ 也随同）趋向于零．

例子来说明一般地数学家怎样在抽象几何里得出结论，怎样一般地可以得出那种与惯常的直观表象无关的结论．第二，我们希望说明，在罗巴切夫斯基几何里可以得出何等多样化的结果．我们再举出一些例子．

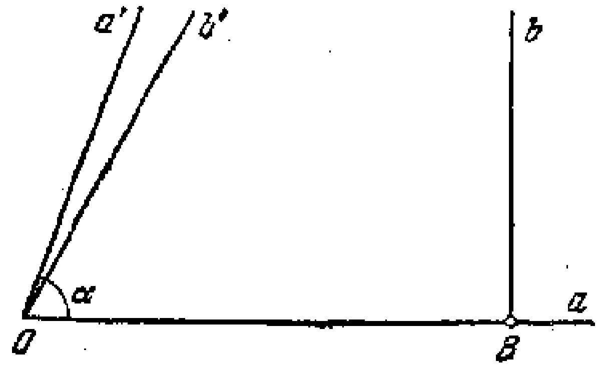


图 8

在罗巴切夫斯基平面上的两条直线，或者相交或者在罗巴切夫斯基意义下平行．那时它们在一侧渐近地逼近，而在另一侧无限地分离，或者它们有公共的垂直线而且在这垂直线两侧都无限地分离．

如果直线 a, b 有公共的垂直线（图 9），则对直线 a 可以引两条垂直线 c, d（在罗巴切夫斯基的意义下），平行于直线 b，而整条直线 b 就处在直线 c, d 之间．

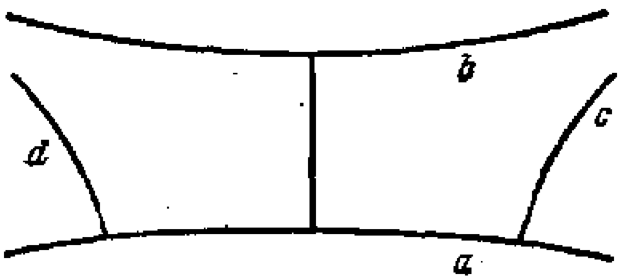


图 9

半径无限增大的圆周的极限不是直线，而是一种曲线，叫做**极限圆**．通过不在一条直线上的三个点，并不总能作一个圆，而能作的或者是圆，或者是极限圆，或者是等距线（即由与一条直线等距离的点组成的曲线）．

三角形的各角之和总小于两直角．假如三角形变大，使得它的所有三条高都无限地增长，则它的三个角全部趋向于零．

不存在面积任意大的三角形．

两个三角形相等，假如它们的三个角都相等．

圆周长度 l 不与半径 r 成正比，而是更迅速地增长（在指数定律的基础上）．那就是说，下列公式成立：

$$l=\pi k(e^{\frac{r}{k}}-e^{-\frac{r}{k}}), \tag{1}$$

这里 k 是依赖于长度单位的常数．因为

$$e^{\frac{r}{k}}=1+\frac{r}{k}+\frac{1}{2}\left(\frac{r}{k}\right)^2+\cdots,\ e^{-\frac{r}{k}}=1-\frac{r}{k}+\frac{1}{2}\left(\frac{r}{k}\right)^2-\cdots,$$

所以我们从公式(1)得到

$$l=2\pi r\left(1+\frac{1}{6}\frac{r^2}{k^2}+\cdots\right). \tag{2}$$

唯有在比值 $\frac{r}{k}$ 很小时才以相当的精确性得出 $l=2\pi r$.

所有这些结论都是下列一些所采用的前提的逻辑推论:"罗巴切夫斯基公理"包容了"通常的"几何的基本的命题.

6. 罗巴切夫斯基几何的特别重要的性质在于:在充分小的区域里它与欧几里得几何很少差异;区域越小,这种差异也越小. 例如,在充分小的三角形里,普通三角学的公式以充分的精确性联系了边和角,并且三角形越小,也就越精确.

公式(2)指出,当半径很小时,圆周长度以很好的精确性与半径成比例.在同样情况下,三角形各角之和与两直角相差很小,等等.

在圆周长度的公式中出现依赖于长度单位的常数 k. 如果半径与 k 相比是很小的,即如果 $\frac{r}{k}$ 很小,则可以从公式(2)看出,长度 l 接近于 $2\pi r$. 一般地说,图形的尺寸与这个常数相比越小,图形的性质越精确地化成欧几里得几何中对应图形的性质[1].

假如 r 表达图形的尺寸(圆的半径,三角形的边,等等),那末比值 $\frac{r}{k}$ 就能表达罗巴切夫斯基几何的图形性质与欧几里得几何的图形性质的差异量.

---

1) 举例说,如果 a, b, c 是一个直角三角形的两条直角边和斜边,则代替毕达哥拉斯定理,下列关系成立:

$$2\left(e^{\frac{c}{k}}+e^{-\frac{c}{k}}\right)=\left(e^{\frac{a}{k}}+e^{-\frac{a}{k}}\right)\left(e^{\frac{b}{k}}+e^{-\frac{b}{k}}\right).$$

展开成级数,我们得出: $c^2+\frac{c^4}{12k^2}+\cdots=a^2+b^2+\frac{a^4+6a^2b^2+b^4}{12k^2}+\cdots$, 因而当 k 很大时我们得到毕达哥拉斯定理 $c^2=a^2+b^2$. 其次,按照罗巴切夫斯基关于平行角 γ 的公式(参看108页的附注), $\operatorname{tg}\frac{\gamma}{2}=e^{-\frac{h}{k}}$. 如果 $\frac{h}{k}$ 很小,即如果平行线很接近,则 $\operatorname{tg}\frac{\gamma}{2}=e^{-\frac{h}{k}}\approx 1$, 因而 $\gamma=90^\circ$. 这样一来,在罗巴切夫斯基平面上相距很近的平行线与欧几里得的平行线相差很小.

由此得出一个重要的结论.

设我们在现实空间中以公里来测量距离. 我们假设这时常数 k 很大, 譬如等于 10^{12}.

于是, 譬如说, 按照圆周的公式(2), 即使它的半径是 100 公里, 圆周长度对半径的比值与 2π 相差还是要小于十万万分之一. 与欧几里得几何的其他关系的差异也有同样的程度. 在一公里的范围内它们已经小到 $\frac{1}{k}$, 即 10^{-12} 的程度,而在一公尺的范围内则还要小到 10^{-15}, 即完全无足轻重的程度. 与欧几里得几何的这样的差异已经无法察觉到了, 因为即使是原子的尺寸也要来得更大些(它们形成 10^{-13} 公里程度的量). 另一方面, 在天文学的尺度方面, $\frac{r}{k}$ 也已经可以说是不太小的了.

因而罗巴切夫斯基假设, 即使在普通的尺度下欧几里得几何以很大的精确性成立了, 但是与它的差异还是可以用天文的观察发觉的. 我们已经指出过, 这个假设本身是实现了, 但是现在在天文的尺度里所发现的那个与欧几里得几何的微小差异, 显得还要更复杂些.

最后, 从所作的论证还推出其他一些重要的结论. 那就是说, 既然常数 k 越大, 与欧几里得几何的差异越小, 那末在极限情形, 当 k 无限变大时, 罗巴切夫斯基几何就变成了欧几里得几何. 这就是说, 欧几里得几何正好是罗巴切夫斯基几何的极限情形. 因而如果在罗巴切夫斯基几何里添上了这个极限情形, 则它也就包括了欧几里得几何, 在这意义下它就显得是更普遍的理论. 由于这个原故, 罗巴切夫斯基把自己的理论命名为"泛几何学", 即普遍的几何学. 理论之间的这种关系在数学和自然科学的发展中经常出现: 新的理论包括了旧的理论作为其极限情形, 这就相当于从部分的结论到更普遍的结论的认识过程.

但是假如不能在欧几里得几何的已经用惯的概念体系中建立罗巴切夫斯基几何的比较简单的现实意义, 那末以上的全部论证和结论还将成为难以理解的智力游戏. 罗巴切夫斯基自己终其一

生并未解决这个问题;他遗留下这一部分给他的后继者,而且直到他的工作第一次发表的将近 40 年以后才得到解决. 我们将在下一节中说明这种解决包括了些什么.

§4. 罗巴切夫斯基几何的现实意义

1. 罗巴切夫斯基几何的直观解释最先在 1868 年得到, 当时意大利几何学家贝尔特拉米指出, 在某种曲面——伪球面——上的内蕴几何符合于罗巴切夫斯基平面片段上的几何. 我们回想到, 曲面的内蕴几何研究的是只由曲面本身上长度的测量决定的那种图形性质的总体. 在图 10 左边画的是所谓曳物线. 这曲线具有这样一些性质, 在它的切线上从切点到与 Oy 轴交点的线段对于所有的点说都是常数. Oy 轴是它的渐近线. 曳物线绕它的渐近线旋转, 我们就得到在图 10 右边所画的曲面, 它叫做伪球面.

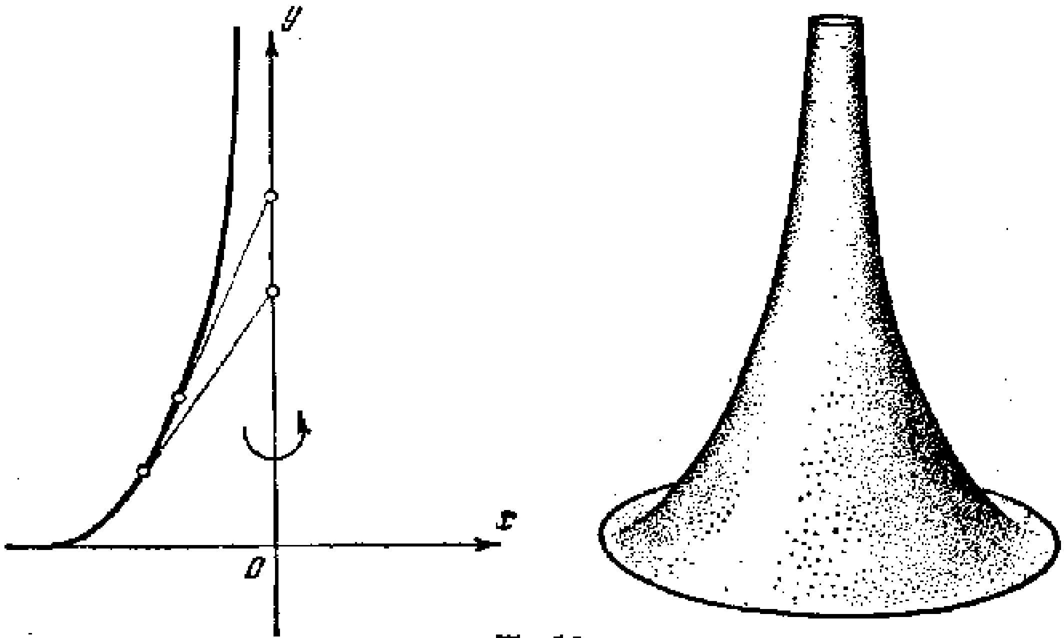


图 10

贝尔特拉米所作的罗巴切夫斯基几何的解释是这样的: 在罗巴切夫斯基平面片段上的所有几何关系与适当的伪球面片段上的几何关系相符合, 只要采用以下的约定. 作为直线段的是曲面上的最短线——测地线. 点之间的距离是作为曲面上连结它们的最短线的长度而决定的. 两个图形被认为是相等的, 假如可以比较

它们的点，使得在对应的点之间的内蕴几何的距离都相等．图形在伪球面上保留它们从内蕴几何观点看来的大小的变动，虽然要随同作弯曲变形，却正好反映了在罗巴切夫斯基平面上的移动．在曲面上像普通那样地测量出来的长度、角度和面积，就相当于罗巴切夫斯基几何里的长度、角度和面积．

贝尔特拉米的解释指出，在这些约定下，对应于罗巴切夫斯基几何的每一个断言（就平面片段来说），就有伪球面的内蕴几何的直接的事实．罗巴切夫斯基几何变得有了完全现实的意义：它正是伪球面上的抽象地叙述的几何学．

应该指出，伪球面的内蕴几何早在贝尔特拉米的发现的 30 年前就已经被闵定格研究过了，闵定格确立了表明其与罗巴切夫斯基几何相符合的那些性质．然而不论是他，或是任何其他的人，都没有注意到这一点，当时罗巴切夫斯基的思想还没有得到充分的传播，贝尔特拉米只是比较了罗巴切夫斯基和闵定格的结论，因而发现了它们的联系．

贝尔特拉米的发现立刻改变了数学家对罗巴切夫斯基几何的关系；它从"虚拟的"变成现实的了[1]．

2. 然而，还要强调一下，在伪球面上实现的并非整个罗巴切夫斯基几何，而是其片段上的几何[2]．因而还没有解决在整个平面

1) 确立罗巴切夫斯基几何的现实意义的历史，实际上还要更复杂些．第一，罗巴切夫斯基自己也已经掌握了运用所谓解析模型的方法来证明它的无矛盾性的办法，但是终其一生还是没有能写出这样的证明．第二，德国数学家黎曼在 1854 年发表了他的理论（参看 § 10），在其中已经包含着贝尔特拉米的结论，但是黎曼并未明白地把它们表达出来；他的报告并未被人理解，而且直到 1868 年他死后才被人发表，那时贝尔特拉米的工作已经出现了．一般地说，从企图证明欧几里得公设到非欧几里得几何的意义完全被人认识这一段罗巴切夫斯基几何的历史，其突出的教育意义在于它指出：真理的发现常常需要经过何等的努力和迂回的道路，然后才能成为简单明了的．

2) 伪球面到处有相等的负的高斯曲率．所有常数负曲率的曲面（至少在微小的片段上）都有同样的内蕴几何，因而都可以用来表达罗巴切夫斯基几何．然而，希尔伯特在 1901 年证明了，任何这种曲面都不能无限制地向所有方向延展而不出现奇异性，因而就不能实现整个罗巴切夫斯基平面．另一方面，年青的荷兰数学家凯伯在 1955 年确定，存在着在其内蕴几何的意义下实现整个罗巴切夫斯基平面的光滑曲面，但是这种曲面虽然是光滑的，却不能作连续的弯曲变形，它们没有确定的曲率．

我们还注意到下面的事实：当罗巴切夫斯基几何在常数负曲率 K 的曲面上实现时，在上节公式中所表述的常数 k 得到下列简单的意义：

$$k^2=-\frac{1}{K}.$$

上甚或在空间中解释罗巴切夫斯基几何的现实意义的问题. 这个问题不多久（在1870年）就被德国数学家克莱因所解决. 让我们来说出他所给的解里包含些什么.

我们在普通的欧几里得平面上取一个圆, 而且只考虑这个圆的内部, 即把它的圆周和圆外的区域排斥在讨论之外. 我们约定把圆的这个内部叫做"平面"——原来它就起着罗巴切夫斯基平面的作用. 我们把圆的弦叫做"直线", 并且根据所采用的约定把弦的处在圆周上的端点除外. 最后, 我们叫做"移动"的是圆的任何这样的变换, 它把圆变成自己而且直线还是变成直线, 即不把圆的弦变成弯曲的. 这种变换的最简单的例子是围绕中心的旋转, 但是要知道, 这种变换是非常广泛的. 以后将会说明这种变换是怎样的变换.

只要引用了这样一些约定的表示, 那末就可以证明, 在圆内部的普通几何的事实就变成了罗巴切夫斯基几何的定理, 而且反过来, 罗巴切夫斯基几何的每一个定理, 就可以解释成圆内部的普通几何的事实.

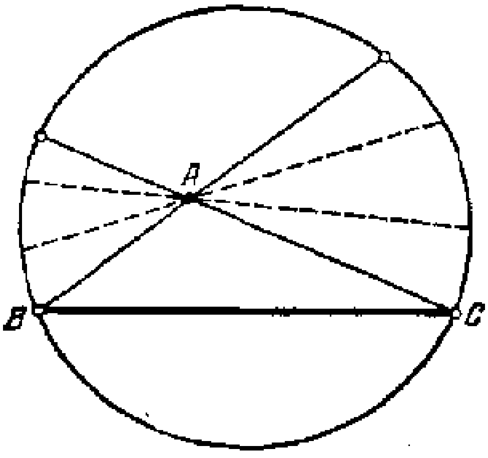


图 11

譬如说, 按照罗巴切夫斯基公理, 通过不在已知直线上的点, 至少可以引两条直线, 不与已知直线相交. 让我们根据所采用的约定, 即把直线换成弦, 把这个公理翻译成普通几何的语言. 那时我们就得到一个断言, 通过圆内不在已知弦上的点, 至少可以引两条弦, 不与已知弦相交. 这个断言的真实性显然可以从图11看出. 因此, 罗巴切夫斯基公理在这时成立了.

其次我们回想起, 在罗巴切夫斯基几何里, 在通过已知点而且不与已知直线相交的直线中间, 有两条是边界, 罗巴切夫斯基就是把这样两条直线叫做已知直线的平行线的. 这表明, 在通过已知点 A 而且不与已知弦 BC 相交的弦中间, 有两条边界弦, 实际上, 这

样两条边界弦就是通过点 B 和点 C 的两条弦. 它们确实与弦 BC 没有公共点，因为我们已经把处在圆周上的点除外了. 这样一来，罗巴切夫斯基的这个定理在这里也成立.

为了进一步把罗巴切夫斯基定理翻译成圆内的普通几何的语言，应该指出在圆内如何测量线段和角，才能使这个测量对应于罗巴切夫斯基几何. 这个测量当然不会与普通的测量相同，因为弦在普通的意义下只有有限的长度，而由弦表达的直线却是无限长的. 这看来似乎会有某种矛盾，但是我们将要看到，这里并无任何矛盾.

首先我们回想起，线段长度的测量是按下列方式进行的. 取某个线段 AB，把它的长度采用作为单位，于是任何别的线段 XY 的长度就可以从它与线段 AB 相比较而决定. 假如线段 XY 还留下小于 AB 的一部分，则就把线段 AB 分成譬如 10 等分（相等的意义是，其中每一分都可以从另一分经过移动而得到）；用这些等分来测量线段 XY 的剩下的部分；然后，假如必要，再把线段 AB 分成 100 等分，依次下去. 结果线段 XY 的长度就以十进小数的形式表示出来，它也可以是不尽小数. 总之，长度的测量通过移置取作单位的线段或其部分来进行，即测量以移动作为基础. 而只要移动已经有定义（在当前的情形下我们把移动定义为把直线变成直线的圆变换），那末随之就知道，哪样的线段算是相等的和应该如何来测量长度. 一句话，移动的定义已经包括了（虽然是以隐蔽的方式）长度测量的法则. 完全同样地，角也可以用取作单位的角来测量. 因此，角的测量法则也包含在移动的定义里.

对应于罗巴切夫斯基几何的长度和角的测量法也可以十分简单地得出，虽然与普通的测量法有实质上的不同. 我们不拟作出它们的推导，因为在我们的论证中这并无原则性的价值[1].

长度的测量法则，使得弦有了无限的长度. 这是因为如果运

---

1) 测量长度的法则如下：设线段 AB 处在弦 CD 上（图12）. 我们用普通的方法测量线段而且组成所谓复合比值 $\frac{CB}{CA}:\frac{DB}{DA}$. 复合比值取对数就被采用作为线段 AB 的长度.

用我们取作移动的变换，线段 AB 变成线段 BB_1，再变成 B_1B_2 等等，则得到的线段在普通的意义下将要变得越来越短(虽然在罗巴切夫斯基几何模型的意义下是相等的；图 12)．点 B_1, B_2, …, B_k, … 向着弦的端点越来越密．但是我们的弦没有端点：按照约定它的端点是除外的，而且它在这意义下已经是“无穷远”的了．在罗巴切夫斯基几何的意义下，点 B_1, B_2, … 并非越来越密，而是趋向无穷远的．运用我们采用作为移动的变换，一个接着一个地截取相等的线段，决不会从圆内达到圆周．

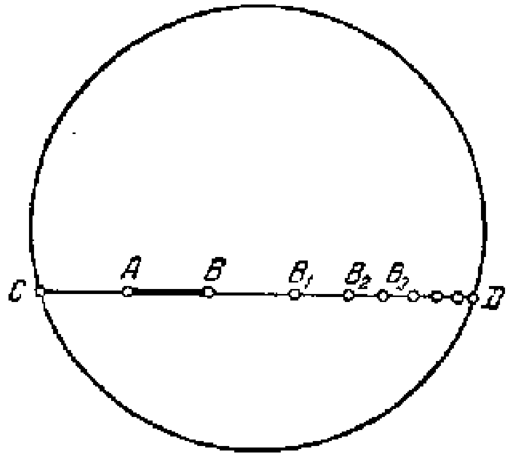


图 12

为了更好地理解在这个模型里如何截取线段，让我们来讨论起着沿直线的平移作用的那种变换．

设在平面上引进了以圆心为原点的直角坐标系．为了确定起见，我们认为我们的圆的半径等于单位，因而它的圆周由方程 $x^2+y^2=1$ 表示，而圆内的点则满足不等式 $x^2+y^2<1$．

我们来讨论由下列公式给定的变换：

$$x'=\frac{x+a}{1+ax},\ y'=\frac{y\sqrt{1-a^2}}{1+ax}, \tag{3}$$

这里 x', y' 是原来有坐标 x, y 的点经过变换而变成的点的坐标，a 是绝对值小于 1 的任何已知数．

假如从公式(3)求出 x 和 y，则我们得到(这很容易验证) x, y 反过来用 x', y' 表示的式子：

$$x=\frac{x'-a}{1-ax'},\ y=\frac{y'\sqrt{1-a^2}}{1-ax'}. \tag{4}$$

变换(3)满足在我们的模型里的“移动”的两个条件：1) 它把圆变成自己；2) 它把直线变成直线．

为了证明第一个性质，实际上就是要肯定：不等式或等式

$x^2+y^2\leqslant 1$ 引出对应的关系 $x'^2+y'^2\leqslant 1$，而且反之亦然. 让我们来证明，当 $x^2+y^2=1$ 时一定有 $x'^2+y'^2=1$，即处在已知圆周上的点还留在圆周上.

我们利用公式(3)而且考虑到 $x^2+y^2=1$，即 $y^2=1-x^2$，来计算 $x'^2+y'^2$：

$$x'^2+y'^2=\frac{(x+a)^2+y^2(1-a^2)}{(1+ax)^2}=\frac{(x+a)^2+(1-x^2)(1-a^2)}{(1+ax)^2}$$

$$=\frac{x^2+2ax+a^2+1-x^2-a^2+a^2x^2}{(1+ax)^2}$$

$$=\frac{1+2ax+a^2x^2}{1+2ax+a^2x^2}=1.$$

因此，当 $x^2+y^2=1$ 时也有 $x'^2+y'^2=1$. 同样地可以验证其余的情形.

要确立变换(3)的第二个性质甚为简单. 事实上，我们知道，每一条直线都由线性方程表示，而且反之，每一个线性方程都表示直线. 设给了直线

$$Ax+By+C=0. \tag{5}$$

经过变换(4)我们得到

$$A\frac{x'-a}{1-ax'}+B\frac{y'\sqrt{1-a^2}}{1-ax'}+C=0,$$

或者去掉公分母有

$$(A-aC)x'+B\sqrt{1-a^2}\,y'+(C-aA)=0.$$

这个方程是线性的，因而它表示直线. 这就是直线(5)经过变换所变成的直线.

我们还注意到，变换(3)把 Ox 轴变成自己，因而只是使点沿着它而挪动. 这是明显的，因为在这条轴上 $y=0$. 而按照公式 (3) 这时还有 $y'=0$. 在 Ox 轴上变换由一个公式给定：

$$x'=\frac{x+a}{1+ax}\ (|a|<1). \tag{3'}$$

在这条直线上，线段 x_1x_2 按照公式(3′)而变成线段 $x_1'x_2'$，而且按照约定，这两个线段算是相等的. 这样也就实行了“线段的截取”.

对于圆心 O，$x=0$，因此 $x'=a$，即圆心在变换 (3′) 下变成有

坐标 $x=a$ 的点.

因为 a 可以任意给定, 只要 $|a|<1$, 所以圆心可以变成沿 Ox 轴的直径上的任何点.

在同一个变换下, 原来处在 A 的点变成有坐标

$$x_1=\frac{a+a}{1+a^2}=\frac{2a}{1+a^2}$$

的点 A_1. 因此线段 OA 在变换 (3) 下变成线段 AA_1, 这样就在由圆的直径表达的"直线"上"截取了"这个线段.

重复同一个变换, 我们可以任意次地继续截取同一个线段. 有坐标 x_n 的点 A_n 变成有坐标

$$x_{n+1}=\frac{x_n+a}{1+ax_n}$$

的点 A_{n+1}. 因而我们得到了有坐标

$$x_0=a,\ x_1=\frac{2a}{1+a^2},\ x_2=\frac{x_1+a}{1+ax_1}=\frac{3a+a^3}{1+3a^2},\ \cdots$$

的点 A, A_1, A_2, $\cdots$. 因为所有线段 A_nA_{n+1} 都从 OA 经过表达移动的变换而得到, 所以它们全都彼此"相等"——这是在由模型表达的罗巴切夫斯基几何意义下的相等. 容易证明, 点 A_n 向着直径的端点而越来越密. 在模型的意义下它们趋向无穷远.

因为 Ox 轴可以具有任何的方向, 所以这样的平移变换可以沿着任何的直径. 把这种变换与绕着圆心的旋转和对任何直径的反射结合起来, 我们可以得到在模型意义下的全部"移动". 这些变换将在下一节作详细的讨论, 在那里将要严密地证明, 在我们的模型里成立的确实是罗巴切夫斯基几何, 特别地, 采用作为移动的变换确实满足几何学里的移动所要服从的全部条件(公理).

让我们再说一遍克莱因提出的罗巴切夫斯基几何的模型是怎样的. 取作平面的是圆的内部; 点是点, 直线是弦(端点除外), 移动是把圆变成自己而且把弦变成弦的变换; 点的位置(点在直线上; 点在另两个点之间)按普通的意义来理解. 测量长度和角(因而还有面积)的法则已经可以从下列事实得出: 定义了移动也就定义了线段和角的相等, 因而也就定义了沿着一个线段截取另一个

线段的运算. 在所有这些约定下，对应于平面上的罗巴切夫斯基几何的每一个定理的是圆内的欧几里得几何的事实，反之，每一个这种的事实都可以用罗巴切夫斯基几何定理的方式来叙述.

完全同样地可以建立空间中的罗巴切夫斯基几何的模型. 取作空间的是某个球的内部(图 13)，直线是弦，平面是圆周在球面上的圆，并且球面本身(因而弦的端点和所说圆的圆周)都除外，最后，移动定义作为把球变成自己而且把弦变成弦的变换.

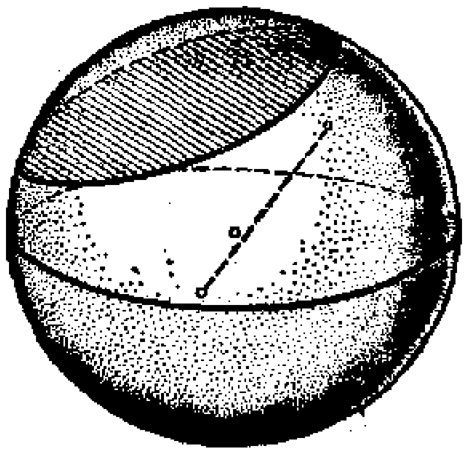


图 13

当罗巴切夫斯基几何的这个模型给出了以后，也就确定了这个几何有简单的现实意义. 罗巴切夫斯基几何是真实的，因为它可以理解成圆内和球内的几何学的独特命题. 这也就证明了它的无矛盾性：它的结论不能引向矛盾，因为它的每一个结论都可以翻译成圆内(如果所说的是空间中的罗巴切夫斯基几何时是在球内)的普通欧几里得几何的语言[1].

3. 在克莱因以后，法国数学家庞加莱给出了罗巴切夫斯基几何的另一种模型，它可以应用到复变数函数论里的一些重要结果的推论上[2]. 因此，在罗巴切夫斯基几何的范围内解决了完全是另一个数学领域的困难问题. 罗巴切夫斯基几何在数学和理论物理里还有许多其他的应用；例如在 1913 年，物理学家瓦里恰克就给出了它在相对论里的应用.

罗巴切夫斯基几何得到了顺利的发展，在其中展开了几何作图的理论、曲线和曲面的一般理论、凸体的理论等等.

---

1) 数学家们通常是这样说的：罗巴切夫斯基几何在欧几里得几何里表达出来，因而它是像欧几里得几何无矛盾一样地无矛盾的.

2) 庞加莱的模型是这样的，仍然取圆的内部作为罗巴切夫斯基平面，但是却把垂直于已知圆周的圆弧算做直线；算做移动的是把圆变成自己的任何保角变换(就是由罗巴切夫斯基几何与保角变换的联系给出它与复变数函数论的联系).

§5. 几何公理，它们利用一定的模型来检验

1. 为了在数学上严密地证明克莱因的模型确实给出了罗巴切夫斯基几何的解释，首先必须正确地表述什么是特别需要证明的. 一个一个地检验罗巴切夫斯基定理是没有道理的；定理很多，并且还是无限地多，因为可以证明新的和更新的定理. 显然，只要能证明在克莱因的模型里罗巴切夫斯基的那些可以导出其余命题的基本的命题成立即可. 但是在这时候必须正确地表述这些基本的命题.

因此，证明罗巴切夫斯基几何无矛盾的问题化成了正确地和完全地表述它的基本命题，即公理的问题. 因为罗巴切夫斯基几何的前提与欧几里得几何的前提相差的只是一个平行公理，所以问题就化成如何正确地和完全地表述欧几里得几何的公理. 欧几里得自己并未作出这样的表述；特别说来，欧几里得就完全没有提出图形的移动或位置的性质的任何定义，虽然他一直利用着它们. 正是随着罗巴切夫斯基几何的发展，同时也随着上一世纪末叶形成的数学原理精确化的一般潮流，欧几里得几何公理的精确化和补充的问题才提高到最高度.

经过许多几何学家们的研究，几何公理的表述问题得到了解决.

一般地说，在采用不同的概念作为基本概念时，就可以不同地选取公理. 我们在这里引进平面上公理的表是这样的，在其中的基本概念是点、直线、移动和这样的一些概念，如像点 X 在直线 a 上；点 B 在点 A 和 C 之间；把点 X 变成点 Y 的移动. (这时候其他的概念就通过它们来定义；举例说，线段就定义为直线上处在两个已知点之间的所有点的集合.)

公理分成五组.

I. 结合公理

1） 通过每两个点有一条直线并且只有一条。

2） 在每条直线上至少有两个点.

3） 至少存在三个点，不在一条直线上。

II. 顺序公理

1） 在直线上的任何三个点中，只有一个点在另外两个点之间.

2） 如果 A, B 是一条直线上的两个点，则在这条直线上至少有一个点 C，使得点 B 在点 A 和 C 之间.

3） 直线把平面分成两个半平面(即直线把平面上的所有不在这直线上的点分成两类，使得同一类的点能被不与直线相交的线段连结起来，而不同类的点则不能这样做).

III. 移动公理

(移动不要理解成个别图形的变换，而要理解成整个平面的变换.)

1） 移动把直线变成直线.

2） 一个接着一个地进行两个移动，相当于进行某一个移动.

3） 设 A, A' 和 a, a' 是两个点和分别从它们引出的两条半直线，α, α' 是分别以延长后的直线 a 和 a' 为界的半平面；存在着唯一的移动把 A 变成 A'，a 变成 a'，α 变成 α'. (直观地说，用平移把点 A 变到 A'，然后用旋转把半直线 a 变到 a'，于是半平面 α 或者与 α' 重合，或者还需要进行绕直线 a 的"翻转".)

IV. 连续公理

设点 X_1, X_2, X_3, … 分布在一条直线上，使得每个在后的点都在其前一个点的右侧，但是同时有着处在所有这些点右侧的

点 A[1]. 那末就存在这样的点 B, 它也在所有的点 X_1, X_2, … 的右侧, 但是有着离它任意近的点 X_n(即不论在点 B 左侧取怎样的点 C, 在线段 CB 上总有点 X_n).

V. 平行公理(欧几里得的)

通过已知点只能引一条直线不与已知直线相交.

以上的公理已经足以构成平面上的欧几里得几何了. 从这些公理确实可以导出中学平面几何课本里的全部定理, 虽然这个推导是非常费事的.

罗巴切夫斯基几何的公理只在平行公理部分才有不同.

V′. 平行公理(罗巴切夫斯基的)

通过直线外的点至少能引两条直线不与已知直线相交.

可以用不多几页篇幅说明, 为什么在公理的表里要有例如公理: "在每条直线上至少有两个点". 谁不知道按照我们关于直线的观念, 在直线上甚至有无限多个点. 但是很奇怪地, 不论是欧几里得还是直到上世纪末叶的任何一个数学家, 都没有想到表述这样的公理: 它是被暗示的. 但是现在情况改变了. 当我们给出几何学的新解释时, 所谓直线已经不是指的普通的直线, 而是某种别的东西: 曲面上的测地线、圆的弦或者别的什么. 所以就发生了精确地和以周全无遗的方式明白表述一切的问题, 这一切是指我们对于用来描述直线的那些对象所需要的. 对于所有其他的概念和公理也是如此.

因此, 就像我们已经说过的, 几何学各种解释的出现乃是它的基本命题精确化的重要媒介之一. 历史情况也是这样: 公理的精确表述是在贝尔特拉米、克莱因和庞加莱的模型以后才出现的.

2. 现在我们来证明, 在克莱因模型里以上列举的全部公理都成立, 只除去欧几里得的平行公理. 正像我们已经在上一节里指出过的 (图 11), 在这里成立的显然不是它, 而是罗巴切夫斯基公

1) "右侧"可以对应地全部换成"左侧".

理．还要检验公理 I—IV．

在模型里的平面是圆的内部(我们认为圆的半径等于 1)．取作点的是点，取作直线的是弦；“点在直线上”和“点在两个点之间”还照普通的意义来理解．由此可知，结合公理、顺序公理和连续公理成立．举例说，第三个顺序公理不过就是弦把圆分成两部分．

还要检验移动公理．移动是定义作为把圆变成自己而且把直线变成直线(即把弦变成弦——译者)的变换的．从这个定义可见，这个变换满足前两个移动公理：满足第一个公理是因为既然把弦当做直线，保留了弦也就表明保留了直线；满足第二个公理是因为假如进行了两个把圆变成圆而且把弦变成弦的变换，则得到的变换还把圆变成圆而且把弦变成弦，即还是一个取作“移动”的变换．

因此，余下的只是第三个移动公理，而它的检验是在这里唯一有困难的．

我们首先注意到，这个公理包含着两个断言．

设 A, A' 是两个点，a, a' 是分别从这两个点引出的半直线，α, α' 是分别以直线 a, a' 为界的两个半平面．

第一个断言是说，把 A 变成 A'，a 变成 a'，α 变成 α' 的移动是存在的．

第二个断言是说，这样的移动只有一个．

可以认为这两个断言都已经在第三章 § 14(参看第一卷)里证明了，但是我们愿意在这里再引出它们的一种证明，这种证明不像在第三章里那样地牵涉到别的更一般的问题．

我们对于模型(即在“半直线”“半平面”“移动”等术语的所采用的理解下)来证明第一个断言的正确性．

我们先假设点 A' 处在圆心．我们选取坐标系，使它们的原点处在圆心，而 Ox 轴通过点 A (图 14)．

在上节中我们讨论过变换

$$x'=\frac{x+a}{1+ax},\ y'=\frac{y\sqrt{1-a^2}}{1+ax}. \qquad (6)$$

在那里证明了它是“移动”(即把已知圆变成自己而且把直线变成

直线).

设 x_0 是点 A 的横坐标，它的纵坐标 $y_0=0$. 因而如果我们取 $a=-x_0$，则根据公式(6)，点 A 就变成有坐标 (O, O) 的点，即变成点 A'.

因为在这时直线变成直线，所以"半直线"(即弦的一段) a 占有某个位置 a'' (图 14). 我们绕着圆心作旋转可以把 a'' 变成 a'. "半平面" α 是以"直线"(弦) a 为界的弓形中的一个. 如果在移动后它已经与 α' 重合，则变换也就结束了；如果不重合，我们再用翻转(即对直径 a' 的反射)把它变成 α'.

这样，把"平移"(6)与旋转(必要时再与反射)合在一起，我们就把 A, a, α 变成 A', a', α'. 但是所有这些"移动"的结果还是"移动"；这个移动把 A, a, α 变成 A', a', α'，即证明了所求移动的存在.

以上我们只谈到当点 A' 处在圆心的特殊情形，现在我们假设它占有任意的位置. 于是按照刚才所证明的，我们可以用某个"移动"把它变成圆心，这个"移动"我们记做 D_1. 这时"半直线" a' 变成某条从圆心引出的"半直线" a''，"半平面" α' 变成某个"半平面"(半圆) α'' (图 15).

已经证明过，我们可以用某个"移动" D_2 把点 A 变成圆心，

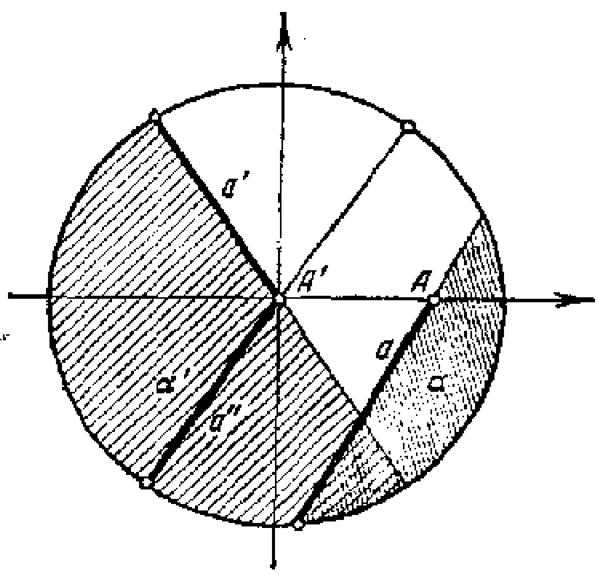


图 14

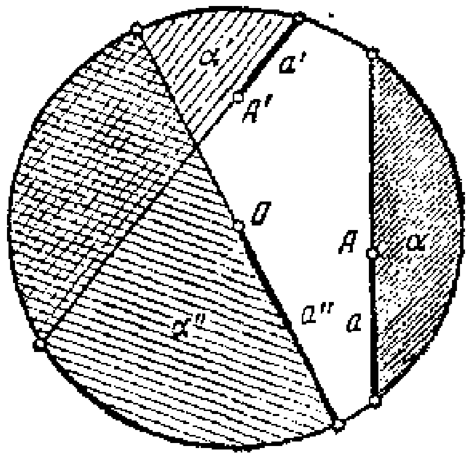


图 15

“半直线” a 变成 a''，“半平面” α 变成 α''. 最后，用与 D_1 相反的“移动”，我们可以把 A' 从圆心变回到原位，同时还把 a''，α'' 变回到原来的位置 a'，α'[1].

因此，“移动” D_2 和与 D_1 相反的“移动”合在一起的结果，我们就把 A, a, α 变成了 A', a', α'. 但是两个“移动”合在一起还是“移动”；因而就确定了对于点 A 和 A' 在圆内的任意位置，存在着把 A, a, α 变成 A', a', α' 的移动. 于是包含在第三个移动公理里的第一个断言也就完全证明了.

现在我们来证明，第二个断言在我们的模型里也成立. 根据这个断言的意义，必须证明下面的事实.

设 A, A' 是圆内的两个点；a, a' 是分别从它们引出的弦段；α, α' 是分别以这两条弦为界的圆弓形. 为了清楚起见，我们把这些点、弦和圆弓形画在两个不同的图上（图 16），虽然它们本来是应该处在一个被讨论的圆内的. 要断定的是，分别把 A 变成 A'，a 变成 a'，α 变成 α' 的“移动”是唯一的，即它完全由这些元素决定.

在作证明时，我们不仅考虑已知圆域的变换，而且考虑整个平面的变换[2]. 按照定义，“移动”把直线变成直线. 具有这种性质的变换叫做**射影变换**. 因此可以说，我们所谓的“移动”是把已知圆

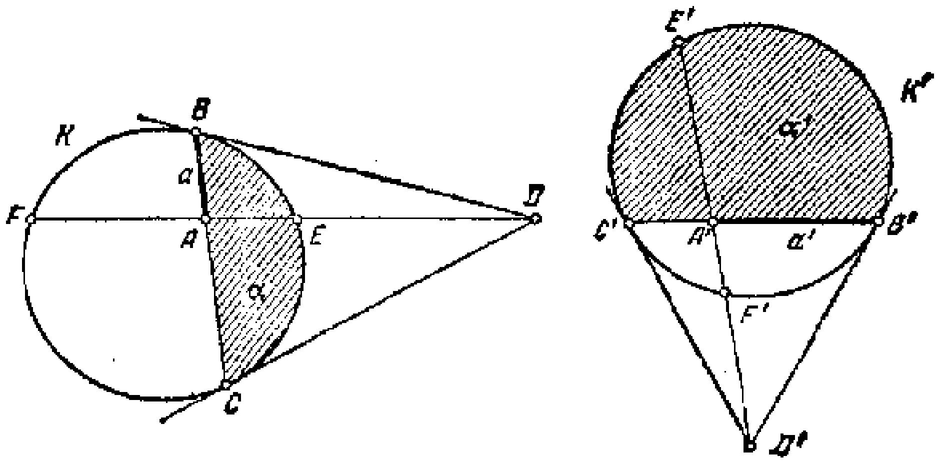


图 16

1) 当 D_1 由公式(3)（§4）表达时，与 D_1 相反的“移动”由公式(4)表达.

2) 可以证明，把圆域变成自己而且把直线变成直线的变换，在保留这些性质下唯一地推广到整个射影平面上，即推广到增添了无穷远点的平面上.

域变成自己的射影变换.（在图 16 上我们画的是把圆域 K 变成圆域 K' 的这种变换. 只需要经过平移就能把圆域 K' 变到 K 上.）

我们曾经在第三章 § 12（第一卷）里讨论过射影变换，这里我们就要利用在那时证明过的下列重要定理：射影变换由没有三个在一条直线上的任何四个点变到哪里而完全决定.

返回到所讨论的"移动". 它把弦段 a 变成 a'，因而就把点 B 变成 B'. 又因为它把弦变成弦，所以它把点 C 变成 C'.

其次，因为一般地说，"移动"把直线变成直线而且把圆域变成自己，而在我们的图上则是把圆域 K 变成圆域 K'，所以在点 B 和 C 处的切线变成在点 B' 和 C' 处的切线. 因而前两条切线的交点 D 变成后两条切线的交点 D'[1]（图 16）.

同时因为点 A 变成 A' 而且直线变成直线，所以直线 AD 变成直线 $A'D'$. 直线 AD 与我们的圆周相交于点 E, F，直线 $A'D'$ 则与它相交于 E', F'.（在图上这些点分别处在圆周 K 和 K' 上.）因为圆域变成自己，所以点 E, F 变成点 E', F'. 设点 E 处在弓形 α 的弧上，E' 处在弓形 α' 的弧上. 那末因为按照条件，α 变成 α'，所以 E 正好变成 E'，对应地 F 就变成 F'.

总之，我们得出结论：在所讨论的"移动"下，圆周上的点 B, C, E, F 变成点 B', C', E', F'. 但是点 B, C, E, F 以及点 B', C', E', F' 显然都没有三个在一条直线上. 因而根据上面所引的定理，把 B, C, E, F 变成 B', C', E', F' 的射影变换是唯一的. 但是"移动"也是射影变换. 因此，我们把 A, a, α 变成 A', a', α' 的"移动"是唯一的，这就是所要证明的.

这样一来，我们证明了在所讨论的模型里欧几里得几何的所有公理（除去平行公理）确实都成立，或者换句话说，在模型里罗巴切夫斯基几何的所有公理都成立. 因此，模型确实实现了罗巴切夫斯基几何. 这个几何就这样地被归结到圆内的欧几里得几何，后者是以术语"直线"和"移动"在模型里所采用的那种约定的理解

1) 假如在点 B 和 C 处的切线平行，则点 D'，是"无穷远点".

而用特别的方式叙述的．顺便提一下，这已经许可在已知的具体模型上展开罗巴切夫斯基几何了，这一点在许多问题里显得是比较方便的．

从几何学基础的逻辑分析的观点看来，上面所作的证明指出：第一，罗巴切夫斯基几何是无矛盾的，第二，关于平行线的公设确实不能从上面列举的其余公理导出．

§6．从欧几里得几何分出的独立的几何理论

1．随着罗巴切夫斯基几何的创立，几何学的原则性的发展还平行地沿着另外的道路行进．在空间的所有几何性质的宝藏中，分出了和进行了由独特的封闭性和稳定性区别出来的各组性质的独立研究．这些按照其方法区分的研究，组成了几何学新的一些部门——关于空间形式的诸科学，就像例如解剖学和生理学组成关于人体的科学的各种部门一样．最初几何学一般说来是不分裂的，它主要研究空间的度量性质——与图形大小的测量有关的性质．只是顺便讨论一下不与测量相关，而与图形的相互位置的量的特征有关的性质，并且早就已经注意到，把这种性质的一部分区别出来的是在图形的形状经过重大的变形和图形的位置经过改变时还保留着的独特的稳定性．

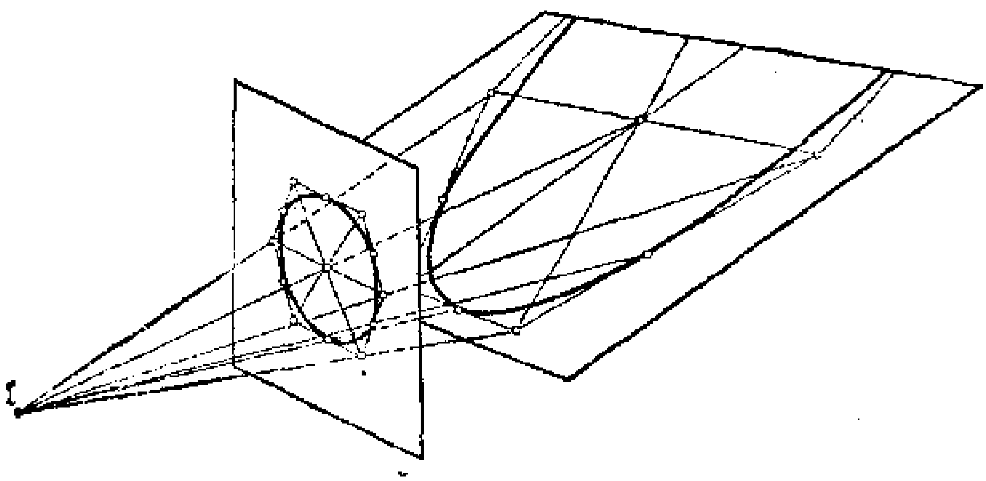


图 17

举例说，我们来察看图象从一个平面到另一个平面的投射(图17)．线段的长度在这时改变了，角改变了，图象的轮廓显然歪曲了．然而举例说，点列处在一条直线上的性质保留着，直线切于某条曲线的性质保留着，等等．

关于投射和射影变换已经在第三章里谈过了．在那里提出了它们与透视——空间图形在平面上的绘象——有显然的联系．透视性质的研究远在欧几里得以前的时代，在古代建筑家的工作中就出现了；研究透视的还有画家丢厄尔和雷翁那陀·达·芬奇，工程师兼数学家戴札格(十七世纪)．最后，在十九世纪初叶，邦色列最先顺次地分出了和研究了在平面或空间的任何射影变换下保留着的几何性质，因而创立了一门独立的科学——射影几何[1]．

本来好像在任意射影变换下保留的性质是不多的和非常原始的，但是却远不是这样．举例说，不久以后将要提到一个定理，这个定理断定，圆内接六边形各对对边延长后的交点处在一条直线上这个事实，对于椭圆、抛物线和双曲线也都成立．这个定理提到的只是射影性质，而后面这些曲线又都可以从圆经过投射而得到．更不显然的是，关于外接六边形的对角线相交于一个点的定理竟是上面所说的定理的独特的同类物；正是在射影几何里揭露了它

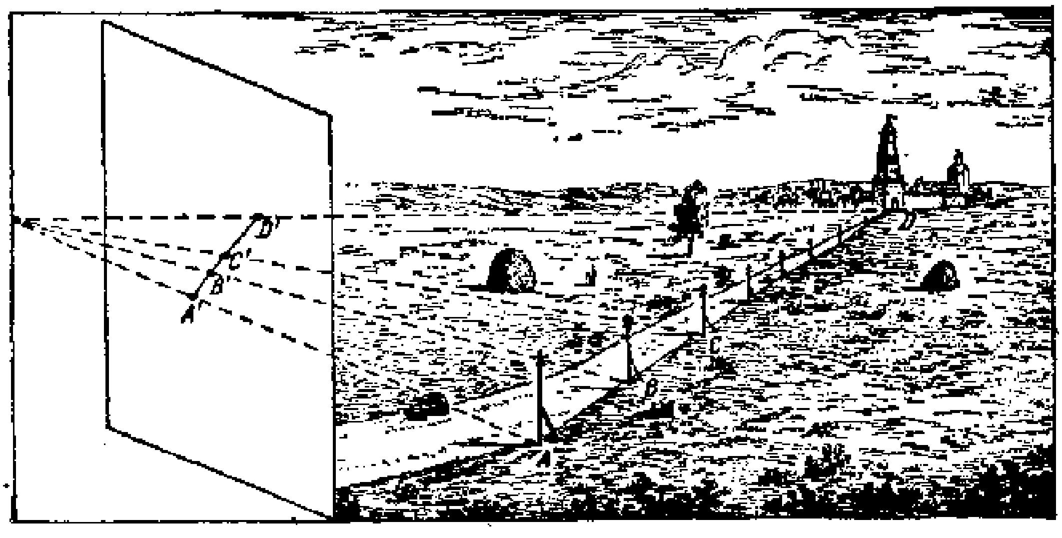


图 18

1) 邦色列是法国的军事工程师，他在 1812 年因被俘虏而逗留在俄国的时候，从事于几何学的研究．他在 1822 年发表了他的工作“论图形的射影性质”．

们的深入的联系．不显然的还有，在射影变换下，不管怎么样改变距离，对于处在一条直线上的每四个点 A, B, C, D（图 18），二重比值 $\frac{AC}{CB}:\frac{AD}{DB}$ 是不变的：

$$\frac{AC}{CB}:\frac{AD}{DB}=\frac{A'C'}{C'B'}:\frac{A'D'}{D'B'}.$$

这引出了在透视里保留的许多关系．举例说，利用这个事实，容易按照伸向远处的道路的照片（图 18），从所看到的电线杆 A', B', C' 的位置，来决定它们到点 D' 的距离．

关于射影几何和它们的结论在航空摄影测量工作里的应用，已经在第三章（第一卷）里谈过了．当然，它的规律可以在建筑学里利用，而且也在构作全景图、布景等等时利用．

射影几何的分出在几何学本身的发展中起着重要的作用．

2. 可以作为独立的几何的另一个例子是仿射几何．在仿射几何里研究的是在任何这种变换下不改变的图形的性质，在那种变换下，每个点的原来位置的笛卡儿坐标 (x, y, z) 和新位置的笛卡儿坐标 (x', y', z') 之间由下列线性方程相联系：

$$x'=a_1x+b_1y+c_1z+d_1,$$
$$y'=a_2x+b_2y+c_2z+d_2,$$
$$z'=a_3x+b_3y+c_3z+d_3$$

（假定行列式

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

不等于零）．

可以证明，任何仿射变换都能化成一个移动（可以加上一个对平面的反射），然后再是空间沿着三个互相垂直的方向的压缩或者伸展．

在任何一个这种变换下，图形保留着很多的性质．直线仍然是直线（一般地说，所有"射影的"性质都保留着）；此外，平行直线仍然平行；保留不变的还有体积的比值，平行平面上或者同一平面

上的两个图形的面积的比值，同一条直线上或者平行直线上的两个线段的长度的比值，等等．很多大家熟知的定理是属于仿射几何的．举例说，三角形的中线相交于一个点，平行四边形的对角线互相平分，椭圆的平行弦的中点处在一条直线上，之类的断言就都是这样的．

与仿射几何有密切联系的有二次曲线(二次曲面)的全部理论．这种曲线就是根据图形的仿射性质而分成椭圆、抛物线和双曲线的：在仿射变换下椭圆正好变成椭圆，既不会变成抛物线，也不会变成双曲线；同样地，抛物线可以变成任何别的抛物线，而不会变成椭圆，等等．

分出了和详细地研究了图形的一般的仿射性质的重要性在于：不太复杂的变换在无穷小的范围内实质上显得是线性的，即是仿射的，而微分计算方法的应用正是与空间的无穷小范围的讨论相联系的．

3．1872 年克莱因在爱耳浪根的大学宣读了现在大家叫做爱耳浪根纲领的演说，他在这篇演说中总结了射影、仿射以及其他“几何”的发展结果，明确地表述了构成这些几何的普遍原则，那就是说，可以考虑空间的一一变换的任何一个群而且研究在这个群的变换下保留着的图形的性质[1]．

从这个观点看来，空间的性质好象是按照它们的深度和稳定性而分出层次．普通的欧几里得几何是经过抛弃现实物体的除几何性质以外的性质而建立的；这里在几何学的特殊部门里，我们已经是在几何学内部再实行一次抽象，抛弃了除决定它们的范围的，我们在已知部分里所关心的几何性质以外的一系列几何性质．

按照所说的克莱因原则可以建立许多的几何．举例说，可以

1) 在这里使用“群”这个字并非单是集合体的意义．在谈到变换群时(参看第二十章)，指的是变换的这种集合，在其中一定包含恒同变换(把所有的点都留在原处的变换)，在其中同时包含每一个非恒同的变换和与它相反的变换(把所有的点都变回到原来的位置的变换)，而且对于这集合中的每两个变换，在这集合中还包含相当于顺次进行这两个变换的结果的变换．

讨论保留任何线之间的角的变换(空间的保角变换)，和研究在这种变换下保留的图形的性质，说到的就是保角几何. 可以讨论不一定是整个空间的变换. 例如在把圆域变成自己而且把弦变成弦的所有变换下讨论圆域的点和弦，而且分出在这些变换下保留的性质，在§4和§5里已经指出过，我们得到的是与罗巴切夫斯基几何相符合的几何.

4. 由这种方式分出的理论的发展，即使在原则性方面(说的已经不是实际的内容)也不止在这里已经谈过的那一些.

譬如我们只考虑图形的仿射性质，我们可以在抛弃所有其他的性质后来设想空间，也可以设想其中的图形只具有我们所关心的性质，因而一般地没有任何其他的性质. 在这个"空间"里，一般说来图形不具有除仿射性质以外的任何其他的性质. 自然可以认为这也是用公理叙述的那种抽象空间的几何，即认为谈的是某些抽象的对象："点"、"直线"、"平面"，它们的性质(这些性质显然要比在欧几里得几何的情形里少些)在某些公理里写出，并且从这些公理导出的推论就对应于普通空间的图形的仿射性质.

这确实可以做到，而且这种抽象的"点"、"直线"和"平面"的集合连同它们的性质的体系就叫做**仿射空间**.

同样地可以设想只具有那种范围的性质的对象的抽象体系，这种性质对应于欧几里得空间的图形的射影性质. (在这时候，公理体系与普通几何的公理的差别显得还要更多些.)

5. 假如进一步深入到几何形象的本质里，那就可以看到，在一系列问题里起作用的还有比射影性质更为深入的性质，这些性质在图形的任何变形下保留着，只要这种变形不造成图形各部分的断裂和折叠. (要使关于这种连续变形的观念精确化，在直觉的描写以外，还可以从分析学中引用我们熟知的连续函数的定义而说，提到的是这种把图形的全部点变到新位置的任意变换，在这种变换下，在新位置的点的笛卡儿坐标由其原来的坐标的连续函数表达，而旧坐标自己也可以表示成新坐标的连续函数.)

在任何这种变换下保留的图形的性质叫做拓扑性质，而研究

它们的科学则叫做拓扑学(参看第十八章).

对于空间图形来说，与图形密切联系的拓扑性质，具有显著的直观性．举例说，几乎是显然的，在平面上每一条从圆周经过连续变形而得到的曲线，总把平面分成在这个任意地弯曲的周界的内部和外部两部分；所以圆周分平面成两部分的性质是拓扑性质．直观上也许是显然的，环面(图19)不可能用任何连续变换变成球面，所以某个曲面许可用连续变换变成环面这个性质，就是使它与很多其他曲面有区别的拓扑性质．

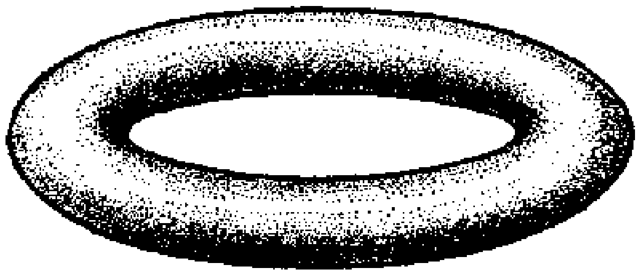


图 19

与连续性有关的论证是与直观有别的，而且常常能很好地说明事情的本质，能把它们变成严密的证明是非常吸引人的，尤其是把同类的例子推广到其他相当复杂的问题上．

举例说，说明代数基本定理的正确性的论证就是如此．所谓代数基本定理是说，每一个方程

$$z^n+a_1z^{n-1}+a_2z^{n-2}+\cdots+a_{n-1}z+a_n=0 \tag{7}$$

至少有一个实数根或者复数根．

设 z 是一个复数平面的点，而 $w=f(z)$ 是与它们对应的另一个复数平面的点 w，并且 $f(z)$ 表示方程(7)的左边．当 z 的绝对值很大时，函数 $f(z)$ 与 z^n 的差别相当地小；而函数 z^n 是非常简单的．特别地，容易验证，如果连续地变动的点在复数平面上画出以坐标原点为中心的圆，则点 $w_1=z^n$ 正好在 w 平面上绕着半径 $|z|^n$ 的同样的圆画 n 次[1]．因而点 $w=f(z)$ 画出 n 个圈，组成某条周线 Γ，相当接近于由点 $w_1=z^n$ 画出的曲线(图20)．

现在如果由点 z 所画的圆周连续地收缩成一个点，则由点

1) 事实上，如果 $z=\rho(\cos\varphi+i\sin\varphi)$，则大家知道[参看第四章(第一卷) §3]，$z^n=\rho^n(\cos\varphi n+i\sin\varphi n)$；所以当数 z 的幅角 φ 从 0 变到 2π 时，数 z^n 的幅角从 0 变到 $2\pi n$，即引向点 z^n 的半径随着点的变动而绕了 n 个整圈．

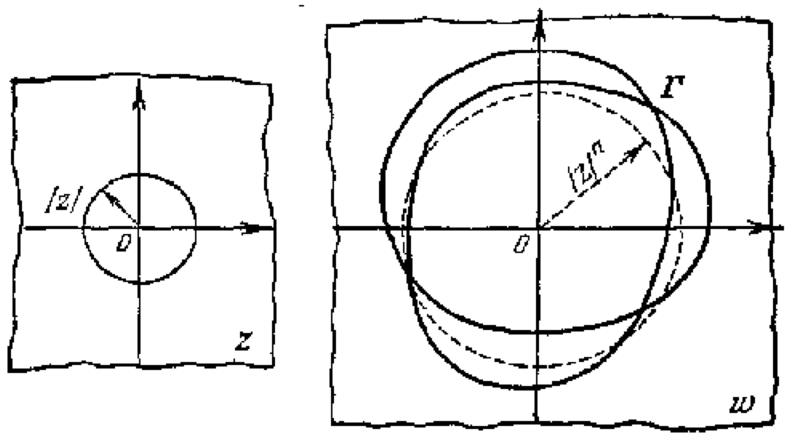


图 20

$w=f(z)$ 所画的绕了 n 圈的周线 Γ 也连续地变形而且收缩成一个点. 但是十分明显地,它在这时假如不越过坐标原点 O,则它就不能收缩成一个点,因为这周线本来是绕着点 O 的.这表示它至少有一度要越过点 O,对于这样的 z 就有 $w=f(z)=0$. 这个 z 就是方程(7)的根.实质上已经可以看出,在所说意义下的根正好有 n 个,因为周线 Γ 的 n 个圈中的每一个在收缩时都要越过点 O.

我们的论证自然要求精确地表示我们在这里利用的周线及其变形的那些拓扑性质.

可以举出在各种不同的、暂时离几何学很远的数学分支里利用拓扑性质的很多例子.

在研究拓扑性质时,在我们之前重又出现设想一般地只具有这种性质的对象的抽象集合的可能性(参看第二十章§7). 这种集合叫做**抽象的拓扑空间**.

这个观点比起拓扑性质作为图形的几何性质的研究来已经是无可比拟地广泛了.拓扑空间可以是各种各样的;譬如说,环面的全部点连同它们的彼此邻接的公共的规律性就组成一个拓扑空间,平面的全部点组成另一个拓扑空间,空间的全部点组成第三个拓扑空间;在第九章(第二卷)§5里讲述复变数函数论时提到过的那些多叶黎曼曲面的全部点组成另一些并且还是不同的拓扑空间.但是最值得注意的是,在远远不像我们关于几何的点的观念的对象之间,常常明白地确立了邻近和邻接的概念. 譬如说,对于某个

活动机械的所有可能的位置，可以明白地指出，什么是“邻近的”位置，什么是一个位置“邻接于”无穷的一串其他的位置，在这些位置中有任意地邻近于已知位置的位置。

我们看到，拓扑空间的概念是极为普遍的，因而我们在§8里将要再一次回到这个概念。

这一节的目的与其说是给读者以关于各种几何的观念，还不如说是企图说明：很多非常具体的问题化成了各组几何性质的分析和研究；它们的研究又引起关于具有这些性质的抽象几何对象的观念的建立，即这些性质的分析以它们的纯净的形状使我们具有关于对应的抽象空间的观念。

在下一节里将要提到也引向构成各种抽象空间的其他的途径。

§7. 多维空间

1. 在新几何思想的发展中的重要的一步是在上一章中已经提到过的多维空间的几何学的创立。它的产生的原因之一是在解决代数和分析的问题时利用几何想法的倾向。解决解析问题的几何途径是以坐标方法为根据的。我们来举一个简单的例子。

假如要求知道不等式 $x^2+y^2<N$ 有多少个整数解。把 x 和 y 看作平面上的直角坐标，我们看到，问题化成这样：在半径 $\sqrt{N}$ 的圆的内部包含着多少个具有整数坐标的点。有整数坐标的点正是铺满平面的边长是1的正方形的顶点(图21)。圆内的这种点的个数近似地等于处在圆内的正方形数，即近似地等于半径 $\sqrt{N}$ 的圆的面积。因此，我们所关心的不等式的解数近似地等于 πN。这时不难证明，可能的相对误差当 $N\to\infty$ 时趋向于零。这个误差的更精确的研究是数论的极为困难的问题，它在不久以前还是深入研究的对象。

为了直接得到从“纯代数”的观点看来远非显然的结果，在所取的例子里把问题翻译成几何的语言显得是足够的。完全同样地

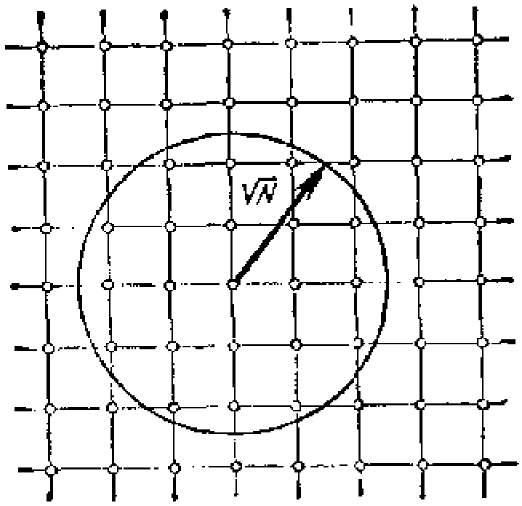


图 21

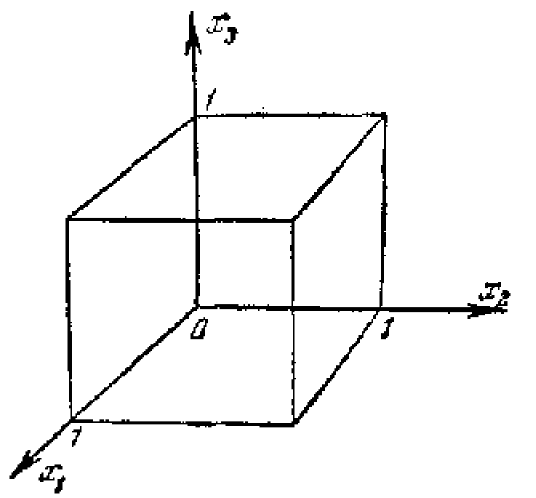


图 22

可以解决关于三个未知数的不等式的同类问题. 但是, 如果未知数多于三个, 则这个方法就不能应用了, 因为我们的空间是三维的, 即点在其中的位置由三个坐标决定. 为了在同样的情形里保留有效的几何同类物, 引出了抽象的 n 维空间的概念, 这种空间的点由 n 个坐标 $x_1, x_2, \cdots, x_n$ 决定. 这时几何学的基本概念是以这样的方式推广的, 使得几何的想法得以解决 n 个变数的问题; 这有力地保证了结论的获得. 这种推广的可能性基于代数规律的统一性, 根据这种统一性, 任意个变数的许多问题的解是完全一样的. 这就许可把在三个变数时起作用的几何想法应用到任意个变数的情形.

2. 四维空间概念的萌芽在拉格朗日的工作中就已经出现了, 他在他的力学方面的工作中形式地把时间当作与三个空间坐标并列的"第四个"坐标. 但是多维几何原理的第一个系统的叙述是在 1844 年由德国数学家格拉斯曼和与他无联系的英国数学家凯利给出的. 他们同时通过了与普通解析几何作形式的类比的途径而进行的. 在现代的叙述中这种类比的大体情况表现如下.

n 维空间的点由 n 个坐标 $x_1, x_2, \cdots, x_n$ 所决定. n 维空间的图形定义成满足这个或那个条件的点的轨迹或集合. 举例说, "n 维立方体"就定义成坐标满足不等式: $a \leqslant x_i \leqslant b\ (i=1, 2, \cdots, n)$ 的点的轨迹. 这里与普通立方体的类比是十分明显的. 在 $n=3$

即三维空间的情形，我们的不等式确实决定一个立方体，这立方体的棱平行于坐标轴而且棱的长度等于 $b-a$（图 22 上画着 $a=0$, $b=1$ 的情形）.

两点之间的距离可以定义成坐标之差的平方和的平方根：

$$d=\sqrt{(x_1'-x_1'')^2+(x_2'-x_2'')^2+\cdots+(x_n'-x_n'')^2}.$$

这是平面上或者三维空间中（即当 $n=2$ 或 3 时）对于距离的熟知公式的直接推广.

现在可以在 n 维空间里定义图形的相等了．两个图形被认为是相等的，假如在它们的点之间可以建立这样的对应，使得每对点之间的距离都等于它们的对应点之间的距离．保留距离的变换可以叫做广义的移动[1]．于是按照与普通欧几里得几何的类比可以说，n 维几何的对象由在广义的移动下保留的图形的性质组成．n 维几何的对象的这个定义是在上一世纪 70 年代确立的，同时也给出了研究它的正确基础．从那时起，n 维几何就成为与欧几里得几何的方向（初等几何、曲线的一般理论等等）类似的所有方向的大量研究的对象.

点之间的距离的概念使我们还能把别的一些几何概念搬到 n 维空间中，例如线段、球、长度、角、体积等等．举例说，n 维球体可以定义成离开已知点不远于已知数 R 的点的集合．因而球体可以解析地用下列不等式给定：

$$(x_1-a_1)^2+\cdots+(x_n-a_n)^2\leqslant R^2,$$

这里 $a_1, \cdots, a_n$ 是球心的坐标．球面则由下列方程给定：

$$(x_1-a_1)^2+\cdots+(x_n-a_n)^2=R^2.$$

线段 AB 可以定义成这样的点 X 的集合，使得从点 X 到 A 和 B 的距离的和等于从 A 到 B 的距离．（线段的长度就是它的端点之间的距离.）

3．我们稍稍详细地来讨论一下各种维数的平面.

在三维空间中这是指一维"平面"——直线和普通的（二维）平

1）推广的不仅在于过渡到 n 个变数，而且在于把对平面的反射也包括在移动之中，因为对平面的反射也不改变点之间的距离.

面. 在 $n>3$ 时的 n 维空间中还要在讨论中引入维数从 3 到 $n-1$ 的多维平面.

大家知道, 在三维空间中, 平面由一个线性方程给定, 直线由两个线性方程给定.

直接推广的途径引向以下的定义: n 维空间里的 k 维平面是指坐标满足 $n-k$ 个线性方程的点的轨迹:

$$\begin{aligned}
&a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n+b_1=0,\\
&a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n+b_2=0,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&a_{n-k,1}x_1+a_{n-k,2}x_2+\cdots+a_{n-k,n}x_n+b_{n-k}=0,
\end{aligned}\tag{8}$$

并且诸方程是相容的和独立的(即其中任何一个都不是其他方程的推论). 这些方程中的每一个都表示一个 $n-1$ 维平面, 而它们全体则共同决定 $n-k$ 个这种平面的公共点.

方程 (8) 相容的意思是说, 一般地有满足这些方程的点, 即已知的 $n-k$ 个 $n-1$ 维平面相交. 没有一个方程是其他方程的推论的意思是说, 这些方程中的任何一个都不能排除. 否则方程组就要化成更少个数的方程, 因而决定的就是更高维数的平面了. 因此, 从几何上来说, 事情成为这样: k 维平面定义成由独立的方程表示的 $n-k$ 个 $n-1$ 维平面的交. 特别地, 如果 $k=1$, 则我们有 $n-1$ 个方程, 它们决定一个"一维平面", 即直线. 因此, k 维平面的所给的定义是解析几何的已知结果的自然的和形式的推广. 这个推广的好处在下列事实中已经可以发觉: 关于线性方程组的结论得到了使得这些结论更清楚的几何解释. 读者可以在第十六章里知道线性代数的问题的这种几何看法.

k 维平面的一个重要性质是它本身可以看作 k 维空间. 譬如说, 三维平面就是普通的三维空间. 这就可以像普通从 n 到 $n+1$ 的论证那样, 把在较低维数的空间中得到的许多结论搬到较高维数的空间中去.

如果方程 (8) 是相容的和独立的, 则在代数学中证明过, 从 n 个变数 x_i 中可以取出 k 个, 使得其余 $n-k$ 个变数可以用它们来

表示[1]．譬如说：

$$x_{k+1}=c_{11}x_1+c_{12}x_2+\cdots+c_{1k}x_k+d_1,$$
$$x_{k+2}=c_{21}x_1+c_{22}x_2+\cdots+c_{2k}x_k+d_2,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$x_n=c_{n-k,1}x_1+c_{n-k,2}x_2+\cdots+c_{n-n,k}x_k+d_k,$$

这里 x_1, x_2, $\cdots$, x_k 可以采取任意的值，而其余的 x_i 则由它们来决定．这说明 k 维平面上的点的位置已经由可以采取任意值的 k 个坐标决定．那就是说，在这意义下 k 维平面有维数 k．

从各种维数的平面的定义可以纯代数地导出下列基本定理．

1）通过不在一个 $k-1$ 维平面上的每 $k+1$ 个点，有而且只有一个 k 维平面．

这里显然与初等几何的熟知定理完全类似．这个定理的证明要运用线性方程组的理论，而且比较复杂，因此我们不叙述它了．

2）如果在 n 维空间里的 l 维和 k 维平面至少有一个公共点，而且这时 $l+k\geqslant n$，则它们相交于维数不低于 $l+k-n$ 的平面．

作为特别情形，由此得出，三维空间中的两个平面假如不重合也不平行，就相交于一条直线($n=3$, $l=2$, $k=2$, $l+k-n=1$)．但是在四维空间中，两个二维平面可以只有一个公共点．举例说，由方程组

$$\left.\begin{matrix}x_1=0\\x_2=0\end{matrix}\right\}\quad 和 \quad \left.\begin{matrix}x_3=0\\x_4=0\end{matrix}\right\}$$

给定的两个二维平面就相交于有坐标 $x_1=0$, $x_2=0$, $x_3=0$, $x_4=0$ 的唯一的点．

所说定理的证明非常简单：l 维平面由 $n-l$ 个方程给定；k 维平面由 $n-k$ 个方程给定；交点的坐标应该同时满足所有 $(n-l)+(n-k)=n-(l+k-n)$ 个方程．如果没有一个方程是其余方程的推论，则按照平面的定义，在相交处的是一个 $l+k-n$ 维平面；

1）在 x_i 中的这 k 个变数，一般说来不能随便选取．举例说，在方程组 $x_1+x_2+x_3=0$, $x_1-x_2-x_3=0$ 中，x_1 的值是唯一确定的：$x_1=0$，因而 x_2 和 x_3 显然就不能选取．但是可以断定，从 x_i 中总可以选取必要的 k 个．

在相反的情形下得到的是更大维数的平面.

对于上面的两个定理还可以补充两个定理.

3) 在每个 k 维平面上至少有 $k+1$ 个不在较低维数的平面上的点. 在 n 维空间中至少有 $n+1$ 个不在任何一个平面上的点.

4) 如果直线与(任何维数的)平面有两个公共点,则它整个处在这个平面上. 一般地, 如果 l 维平面与 k 维平面有 $l+1$ 个公共点不在 $l-1$ 维平面上,则它就整个处在这个 k 维平面上.

我们注意到, n 维几何可以从公理出发而建立, 这些公理是 §5 里表述的公理的推广. 在这种观点下,上面所说的四个定理就采用作为结合公理. 这正好表明公理的概念是相对的: 同一个断言在一种理论结构下是定理,在另一种结构下可以是公理.

4. 我们得到了关于多维空间的数学概念的一般观念. 为了说明这个概念的现实的物理意义,我们重新回到图象表示的问题. 例如我们希望表示气体压力对容积的相关性. 我们在平面上取坐标轴,而且在一条轴上表出容积 v,而在另一条轴上表出压力 p. 在这种约定下,压力对容积的相关性就由某条曲线表出(根据熟知的波义耳-马利奥特定理,在一定温度下对于理想气体这曲线是双曲线). 但是如果我们有更复杂的物理体系,它的情况已经不是由两个条件(例如在气体情形的容积和压力)给定, 而是由譬如五个条件给定,则它的状况的图象表示对应地就引向五维空间的观念了.

例如设谈的是三种金属的合金或者三种气体的混合. 混合物的情形由四个条件决定: 温度 T, 压力 p 和两种气体的百分含量 c_1, c_2 (第三个气体的百分含量将由于全部百分含量等于 100% 而决定,因而 $c_3=100-c_1-c_2$). 因此这种混合物的状况由四个条件决定. 它的图象表示不是需要联合几个图表, 就是必须把这个状况设想成具有四个坐标 T, p, c_1, c_2 的四维空间的点. 这种想法确实在化学里利用了; 美国学者吉卜斯和苏联物理化学家库尔纳可夫院士的学派就研究了在化学问题里运用多维几何的方法. 在这里引入多维空间表明了保留方便的几何类比的倾向和从简单的图象表示方法出发的想法.

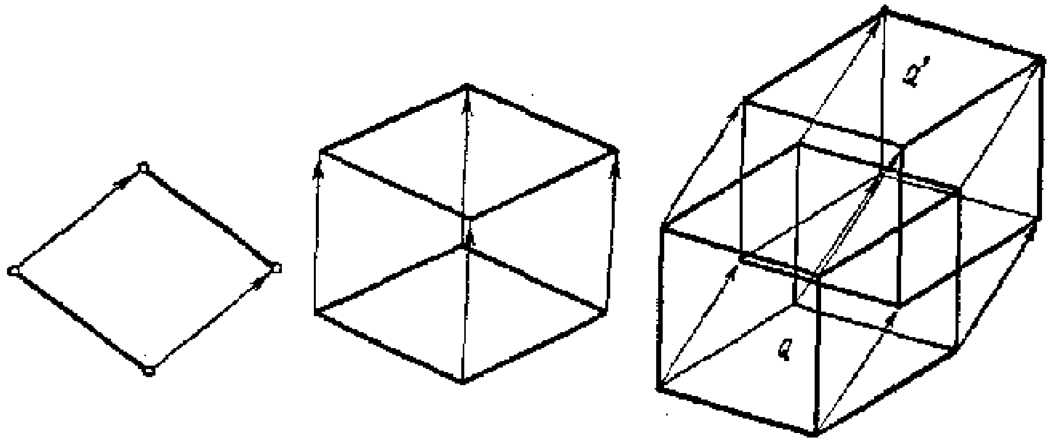
我们再从几何领域里引一个例子．球体由四个条件给定：球心的三个坐标和半径，所以球体可以看作四维空间里的点．因而将近四十年前由某些数学家所建立的球体的特殊几何可以看作是某种四维几何．

以上都说明了引入多维空间概念的一般的现实基础．如果某个图形或某个体系的状况等由 n 个条件给定，则这个图形、这种状况等就可以理解成某个 n 维空间的点．这种想法的好处就象普通的图象的好处一样：它们提供了在研究所讨论的现象时应用熟知的几何类比和方法的可能性．

因此在多维空间的数学概念里没有任何神秘之处．它只不过是数学家们所拟定的某个抽象概念，以便用几何的语言来叙述那些不能作普通意义下的简单几何表示的事物．这个抽象概念有十分现实的基础，它反映了现实而且是为科学的需要而产生的，因而并非无聊的概念游戏．它反映出的事实是：存在着由若干个条件决定的事物（例如球体或三种气体的混合物），因而所有这种事物的集合是多维的．在已知情形里的维数正是这些条件数．就象点在空间中移动时它的三个坐标都改变一样，球体在移动、扩大和缩小时改变它的四个"坐标"，即改变决定它的四个量．

在以下几段里我们还要谈到多维几何．而本段的要点在于说明它是叙述现实事物和现象的数学方法．关于我们的现实空间处在那种四维空间之中的观念——被某些小说家和心灵论者所利用的观念，是与四维空间的数学概念毫无关系的．如果在这里可以说到与科学的关系，那也只不过是科学概念的荒唐的曲解罢了．

5．上面已经说过，多维空间的几何学最初是通过把普通的解析几何形式地推广到任意个变数而建立的．但是事情的这种处理，数学家们是不能完全满意的．要知道目的不仅在于推广几何概念，而且在于推广几何的研究方法．因而不依赖于解析工具而给出 n 维几何的纯几何的叙述就有了其重要性．这最先是由瑞士数学家史雷夫里在 1852 年做到的，他在他的工作中讨论了多维空间的正多面体的问题．当然，史雷夫里的工作没能得到同时代人的评价，



(a) (b) (c)

图 23

因为要理解它在某种程度上必须提高到几何学的抽象见解的高度上．唯有数学的进一步发展才把这个问题完全弄清楚了，而且详尽地说明了解析的和几何的处理的相互关系．由于没有可能深入这个问题，我们现在只举出 n 维几何的几何叙述的几个例子．我们来讨论 n 维立方体的几何定义．在平面上让线段向着垂直于它的方向移动一段等于它的长度的距离，我们就画出了正方形，即二维立方体（图 23 a）．完全同样地，让正方形向着垂直于其平面的方向移动一段等于其边长的距离，我们就画出一个三维立方体（图 23 b）．为了得到四维立方体，可以应用同样的作图：在四维空间里取一个三维平面和在其中的一个三维立方体，我们让这个立方体向着垂直于这个三维平面的方向移动一段等于棱长的距离（按照定义，直线垂直于 k 维平面，假如它垂直于处在这个平面上的每一条直线），这个作图假定如图 23 c 所画．图上画着两个三维立方体 Q 和 Q'——在最初位置和最终位置的已知立方体．连结这两个立方体的对应顶点的线画成这样的线段，立方体的顶点就是沿着它们而移动的．我们看到，四维立方体一共有 16 个顶点：立方体 Q 的八个顶点和立方体 Q' 的八个顶点．其次，它有 32 条棱：变动的三维立方体在最初位置 Q 的 12 条棱和它在最终位置 Q' 的 12 条棱，还有 8 条"侧面的"棱．它有 8 个三维面，这些面本身是立方体．在三维立方体移动时它的每个面都画出一个三维立方体，

因而得到6个立方体——四维立方体的侧面，此外它还有两个面：对应于变动的立方体的最初位置和最终位置的"前部的"面和"后部的"面. 最后，四维立方体还有总数24个二维正方形面：立方体Q和Q'各有6个，还有12个是由立方体Q的棱在移动时画出的.

总之，四维立方体有8个三维面，24个二维面，32个一维面(32条棱)，最后还有16个顶点；每个面都是对应维数的"立方体"：三维立方体，正方形，线段，顶点(它可以看做零维立方体).

同样地，让四维立方体"向第五维"移动，我们得到五维立方体，而且只要重复这个作图，就可以作出任何维数的立方体. n维立方体的所有的面本身是较低维数的立方体：$n-1$维的、$n-2$维的等等，最后是一维的，即是棱. n维立方体在每一维数有多少个面这个问题是一个有趣的和不难的问题. 容易肯定，它有$2n$个$n-1$维面和2^n个顶点. 试问它有多少条棱?

我们再来讨论n维空间的一个多面体. 在平面上的最简单的多边形是三角形——它有最小可能的顶点数. 为了得到有最小顶点数的多面体，只要取不在三角形平面上的一个点，而且用线段连结它与这个三角形的每个点. 得出的线段填满一个三棱锥(图24). 为了得到四维空间里的最简单的多面体，我们这样地进行讨论. 取某个三维平面和在其中的一个四面体T. 然后取不在已知三维平面中的一个点，我们用线段连结它与四面体T的每个点. 在图24的最右边假定画着这个作图. 连结点O与四面体T的点的每一个线段，与四面体都没有别的公共点，因为否则它就将完全处在包含T的三维空间中了. 所有这种线段是"引向第四维的". 它们填满了一个四维多面体——所谓四维单纯形. 它的三维面是四面体：一个在底部，还有4个是在底的三个二维面上的侧面，一共是5个面. 它的二维面是三角形，一共有10个：四个在底部，还有六个侧面. 最后，它有10条棱和5个顶点.

重复这种作图，我们可以对于任何维数n得到最简单的n维多面体——所谓n维单纯形. 从作图看出它有$n+1$个顶点. 可以肯定，它的所有的面都是较低维数的单纯形：$n-1$维的，$n-2$

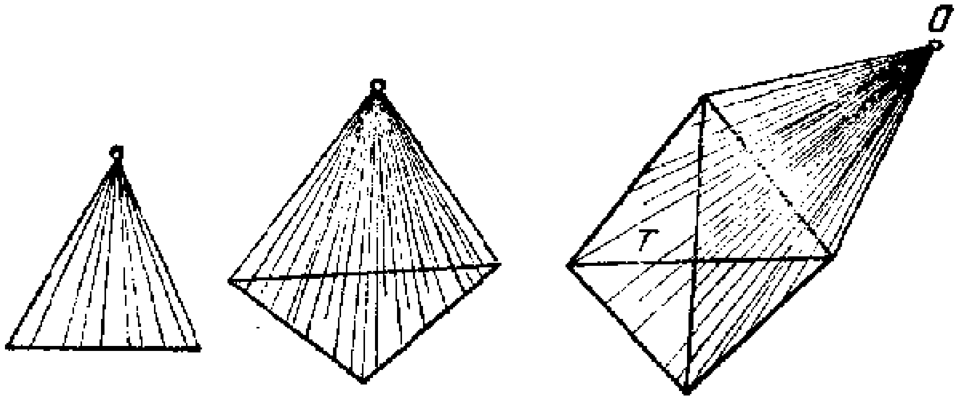


图 24

维的，等等[1]。

棱柱和棱锥的概念也容易推广．如果我们让多边形从平面向第三维作平行移动，则它就画出一个棱柱．同样地，让三维多面体向第四维移动，我们就得到四维棱柱(假定这是如图 25 所画的)．四维立方体显然是四维棱柱的特殊情形．

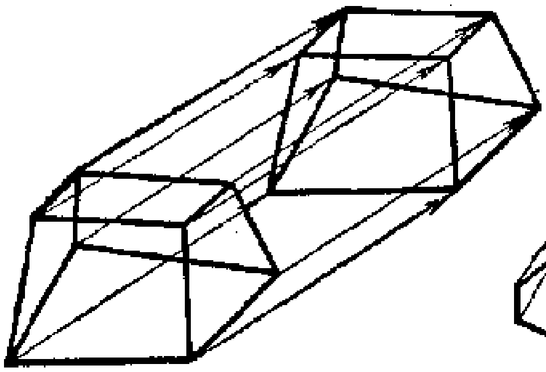


图 25

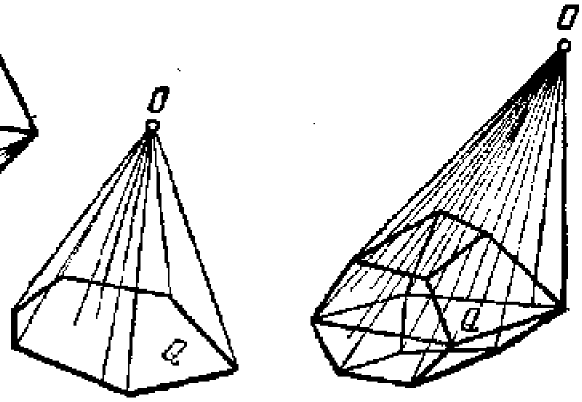


图 26

棱锥用下列方法作出．取多边形 Q 和不在多边形平面上的点 O．用线段连结多边形 Q 的每个点和点 O，这些线段填满一个具有底部 Q 的棱锥(图 26)．同样地，如果在四维空间中给了三维多面体 Q 和不与它在同一个三维平面上的点 O，则连结多面体 Q 的点

1) 单纯形的任何 m 个顶点决定"张在它们之上"的 $m-1$ 维单纯形：已知 n 维单纯形的 $m-1$ 维面．所以 n 维单纯形的 $m-1$ 维面的个数就等于从它的所有 $n+1$ 个顶点中取 m 个的组合数，即等于

$$C_{n+1}^{m}=\frac{(n+1)!}{m!(n-m+1)!}.$$

与点 O 的线段填满一个具有底部 Q 的四维棱锥. 四维单纯形不是别的, 正是以四面体为底的棱锥.

完全同样地, 从 $n-1$ 维多面体 Q 出发可以定义 n 维棱柱和 n 维棱锥.

一般地说, n 维多面体是 n 维空间的一部分, 它以有限个 $n-1$ 维平面的小块为界; k 维多面体是 k 维平面的一部分, 它以有限个 $k-1$ 维平面的小块为界. 多面体的面本身是较低维数的多面体.

n 维多面体的理论是普通的三维多面体的理论的富有具体内容的推广. 在一系列的情形里, 关于三维多面体的定理可以没有特别困难地推广到任何维数的情形, 但是也有一些问题, 它对于 n 维多面体是非常难以解决的. 其中可以提出沃洛诺依 (1868—1908) 的深入研究, 那是在与数论问题的联系中产生的; 苏联几何学家们继续了这些研究. 这些问题中的一个——所谓"沃洛诺依问题"——一直还没有完全解决[1].

在不同维数的空间之间有着本质的差别, 这可以举正多面体为例. 在平面上, 正多边形可以有任意的边数. 换句话说, 存在着无穷多种不同形状的正"二维多面体". 三维正多面体一共有五种形状: 四面体, 立方体, 八面体, 十二面体和二十面体. 在四维空间里有六种正多面体, 但是在任何更大维数的空间里它们一共只有三种了. 这三种是: 1) 四面体的同类物——正 n 维单纯形, 即各棱都相等的单纯形; 2) n 维立方体; 3) 八面体的同类物, 它可以用下列方法作出: 取立方体各面的中心作为这个多面体的顶点, 而它就张在这些顶点上. 在三维空间的情形, 在图 27 上画着这个作图. 我们看到, 对于正多面体说, 二维、三维和四维空间占有独特地位.

---

1) 指的是寻求这样的凸多面体, 把它们平行地沿着整个面放置而可以铺满整个空间. 在三维空间的情形, 这个问题由于结晶学的需要而由费得洛夫提出和解决了; 沃洛诺依和它的后继者们把同一个问题推广到 n 维空间的情形, 但是只是对于二维、三维和四维空间, 这个问题才得到最终的解决.

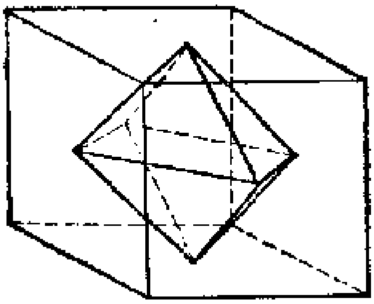


图 27

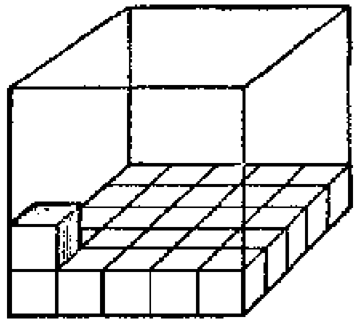


图 28

6. 我们再来讨论 n 维空间里的体积的问题. n 维立体的体积可以像在普通几何里那样地决定. 体积就是用来比较图形的数值特征, 并且在体积上要求两点: 1) 相等的立体有相等的体积, 即体积在图形作为刚体而移动时不改变; 2) 当一个立体从两个立体相加而成时, 它的体积等于它们的体积之和. 采用作为体积单位的是棱长等于 1 的立方体. 这以后就可以确定棱长为 a 的立方体的体积等于 a^n. 像在平面上和在三维空间中那样, 可以做到用一层一层的立方体来铺满一个立方体(图 28). 因为现在是向着 n 个方向放置立方体, 所以总数是 a^n.

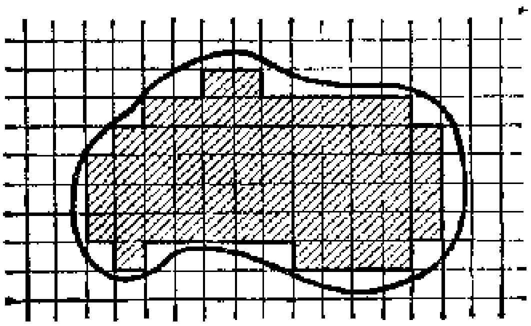


图 29

为了决定任何 n 维立体的体积, 人们把它近似地换成由非常小的 n 维立方体堆成的立体, 就像在图 29 上把平面图形换成由正方形组成的图形一样. 立体的体积就定义为这种阶梯状立体当组成它们的立方体无限制地变小时的极限.

完全同样地可以定义处在某个 k 维平面上的 k 维图形的 k 维体积. 从体积的定义容易得出它的一个重要性质: 当立体相似地变大时, 如果它的所有线性尺度都扩大到 λ 倍, 则 k 维体积扩大到 λ^k 倍.

如果立体剖分成平行的各层, 则它的体积就是各层体积的和

$$V=\sum V_i.$$

每一层的体积可以近似地表示成它的高 Δh_i 乘上对应截面的 $n-1$ 维体积（"面积"）S_i. 结果整个立体的体积就可以近似地表示成一个和式

$$V\approx\sum S_i\Delta h_i.$$

在所有 $\Delta h_i\to 0$ 时过渡到极限，我们得到用积分表示的体积

$$V=\int_0^H S(h)\,dh, \tag{9}$$

这里 H 是立体在垂直于所作截面的方向上的伸展度.

这完全类似于在三维空间中计算体积. 举例说，棱柱的所有截面都相等，因此截面的"面积" S 与 h 无关. 所以对于棱柱有 $V=SH$，即棱柱的体积等于底"面积"和高的乘积. 我们再来决定 n 维棱锥的体积. 设给了具有高度 H 和底面积 S 的棱锥. 我们用平行于底面而且与顶点相距 h 的平面与它相截. 截下的是高 h 的一个棱锥. 我们把它的底面积记做 $s(h)$. 这个较小的棱锥显然与原来的相似：它的所有尺度缩小的程度与 h 小于 H 的程度相同，即缩小到 $\frac{h}{H}$. 因而它的底面的 $n-1$ 维体积（即"面积"）就等于

$$S(h)=\left(\frac{h}{H}\right)^{n-1}S,$$

因为一个 $n-1$ 维图形在线性尺度改变 λ 倍时，它的体积要乘上 λ^{n-1}.

根据公式(9)，整个棱锥的体积等于

$$V=\int_0^H S(h)\,dh,$$

于是

$$V=\int_0^H S\left(\frac{h}{H}\right)^{n-1}dh=\frac{S}{H^{n-1}}\int_0^H h^{n-1}\,dh$$
$$=\frac{S}{H^{n-1}}\cdot\frac{H^n}{n}=\frac{1}{n}SH,$$

即 n 维棱锥的体积等于底"面积"（$n-1$ 维体积）乘高的乘积的 $\frac{1}{n}$. 当 n 等于 2 或 3 时，我们就得到作为特别情形的熟知结果：三角形

的面积(二维体积)等于底乘高的乘积的一半，三维棱锥的体积等于底面积乘高的乘积的三分之一。

球体可以近似地看做由具有公共顶点在球心处的很细的棱锥组成.这些棱锥的高都等于半径 R，而它们的底面积 σ_i 加起来则近似地等于整个球面 S. 因为每个棱锥的体积等于 $\frac{1}{n}R\sigma_i$，所以在把这些体积加起来后，我们得到球体的体积

$$V\approx\frac{1}{n}R\sum\sigma_i\approx\frac{1}{n}RS.$$

在极限情形这给出精确的公式：$V=\frac{1}{n}RS$，即球体的体积等于半径乘球面的乘积的 $\frac{1}{n}$. 当 n 等于 2 或 3 时，这公式是大家熟知的[1].

让我们指出球的一个重要性质，一般说来，这在 n 维空间的情形可以像在三维空间的情形完全一样地证明：在体积一定的所有立体中间，球体(也唯有球体)有最小的表面体.

7. 我们在上面只谈到 n 维空间的初等几何，但是在其中也可以展开"高等"几何，例如曲线和曲面的一般理论. 在 n 维空间里的曲面可以有各种的维数：一维"曲面"即曲线，二维曲面，三维曲面，…，最后是 $n-1$ 维曲面. 曲线可以定义为这种点的轨迹，点的坐标连续地依赖于某个变数——参数 t

$$x_1=x_1(t),\ x_2=x_2(t),\ \cdots,\ x_n=x_n(t).$$

曲线正是点在 n 维空间中随 t 的改变而移动的点的轨线. 如果我们的空间是为了表示某个物理体系的情况的，就像在第 4 段里所说的那样，则曲线将表出情况的连续顺序，或者是与参数 t(例如时间)有关的情况改变过程. 这就推广了利用曲线表达情况改变过程的普通的图象表示.

n 维空间里的曲线在每个点处不仅有切线("一维密切平面")，

---

1) 球体的体积的计算也可以应用公式(9)来实行；n 维球体的截面是 $n-1$ 维球体，因而 n 维球体的体积可以通过从 $n-1$ 到 n 的过程来计算.

而且还有从 2 到 $n-1$ 的所有各维密切平面. 这些平面中每一个的旋转速度对曲线长度的增长速度的比值给出对应的曲率. 因此, 曲线有从一维到 $n-1$ 维的 $n-1$ 个密切平面和对应的 $n-1$ 个曲率. 在 n 维空间里的微分几何显得比三维空间大为复杂, 所以我们没有可能在这里谈曲面的理论.

到此为止, 谈的都是直接推广普通的欧几里得几何的 n 维几何. 但是我们已经知道, 除欧几里得几何外, 还存在罗巴切夫斯基几何, 射影几何和其他的几何, 这些几何也很容易推广到任意的维数.

§8. 几何对象的推广

1. 在上一节里说到关于 n 维空间的现实意义时, 我们已经密切地接近了关于推广几何对象的问题, 接近了关于数学中一般的空间概念的问题. 但是在给出相应的普遍定义之前, 让我们再来讨论一些例子.

经验表明, 正常人的视觉是三色的(这是罗蒙诺索夫最先指出的), 即每一种颜色感觉——颜色 $Ц$——是三种基本的感觉的合成: 红色 K, 青色 $З$ 和蓝色 C, 它们分别具有一定的强度[1]. 把以某种单位计算的这些强度记做 x, y, z, 就可以写成 $Ц=xK+yЗ+zC$. 就像点在空间中可以从上到下、从右到左、从前到后地移动那样, 颜色的感觉——颜色 $Ц$——也可以连续地随着组成它的红色、青色和蓝色部分的改善而向着三个方向改变. 所以按照类比就可以说, 所有颜色的集合是"三维的颜色空间". 强度 x, y, z 起着点——颜色 $Ц$ 的坐标的作用. (与普通坐标的一个重要的不同在于强度不能是负的. 当 $x=y=z=0$ 时我们得到完全的黑色, 它

1) 谈到的是颜色的感觉而不是光线. 颜色的感觉也是客观的现象——光线的作用. 同一个色感可以由不同的光波引起. 举例说, 青色不仅可以从光谱上纯粹的青光得到, 还可以从红光和蓝光的混合而得到. 另一方面, 患有"色盲"的人(色盲者)只有两种基本的色感; 而在"完全色盲"的情形, 就会只有一种基本的色感, 这是极为稀少的.

对应于完全没有光线.)

颜色的连续改变可以表示成"颜色空间"中的曲线,譬如虹霓的颜色就组成一条这种曲线;颜色曲线还表示由均匀色彩的对象在光线亮度连续改变下所引起的一系列的色感.在这情形里改变的只是色感的强度,它的"色度"保留不变.

然后,如果给了两种颜色,例如红色 K 和白色 B,则把它们按不同的比例混合后[1],我们就得到从 K 到 B 的颜色的连续序列,它可以叫做线段 KB.关于玫瑰色处在红色和白色之间的观念是有明白的意义的.

因此,在"颜色空间"里产生了关于最简单的几何图形和关系的概念."点"是颜色,"线段" AB 是由颜色 A 和 B 混合而得到的各种颜色的集合;"点 D 在线段 AB 上"则表示颜色 D 是颜色 A 和 B 的混合.三种颜色的混合给出平面块——"颜色三角形".这一切都可以利用颜色的坐标 x, y, z 而写成解析式子,并且给定颜色直线和颜色平面的公式完全类似于普通解析几何里的公式[2].

在颜色空间里,欧几里得几何的关于点和线段的位置的关系成立,关于这些关系的学问组成了仿射几何,因而可以说在所有可能的色感的集合里实现了仿射几何.(当然这并不完全正确,因为已经说过,颜色的坐标 x, y, z 不能是负的,所以颜色空间只对应于这样一部分空间,那里在已知坐标系里点的所有坐标都是正数或零.)

其次,我们有着关于颜色的差异程度的观念.譬如说,明显地,淡玫瑰色比深玫瑰色更接近于白色,紫红色比蓝色更接近于红色,等等.因此,我们有了关于颜色之间的距离作为它们的差异程度的定性的概念.这种定性的概念可以给以定量的测度.然而像在欧几里得几何里那样按照公式

1) 在色感保留不变的条件下以不同的比例混合极细的颜色粉,就可以得到这种混合色.

2) 举例说,如果颜色 Π_0 和 Π_1 由强度——坐标 x_0, y_0, z_0 和 x_1, y_1, z_1 决定,则处在 Π_0 第 Π_1 之间的颜色 Π 就有坐标 $x=(1-t)x_0+tx_1$, $y=(1-t)y_0+ty_1$, $z=(1-t)\cdot z_0+tz_1$,这里 t 是颜色 Π_1 在形成颜色 Π 的混合物里的成分,$1-t$ 是颜色 Π_0 的成分.

$$r=\sqrt{(x_0-x_1)^2+(y_0-y_1)^2+(z_0-z_1)^2}$$

来决定颜色之间的距离是不自然的. 这样定义的距离不对应于现实的色感;在这种定义下将会得到一系列这种情况,与已知颜色有不同的差异程度的两种颜色,处在离它相同距离的地方. 距离的定义应该反映出色感之间的现实关系.

因此,在颜色空间中要引出独特的测量距离法. 这可以用下列方式做到.

当颜色连续地改变时,人们并不立刻感觉到这个改变,而要等到它达到某种程度,达到所谓区别阶段时. 因此可以认为,恰好与已知颜色距离一个区别阶段的所有颜色都与已知颜色等距离. 于是,我们自然就得出这样一个结论:在任何两种颜色之间的距离应该用正好可以插入它们之间的区别阶段的最小数来量出. 颜色曲线的长度就用在这曲线上共有多少这种阶段来量出. 两种颜色之间的距离由连结它们的最短线的长度决定. 这就像曲面上的两点之间的距离用连结它们的最短线的长度量出一样.

因此,在颜色空间里的长度和距离是用很小的(如同无限小的)一步步来量出的.

结果在颜色空间里可以定义某种独特的非欧几里得几何. 这个几何有十分现实的意义:它以几何的语言描述了颜色集合的性质,即光线的刺激对眼睛的作用的性质.

关于颜色空间的概念在将近一百年前就产生了. 研究了这个空间的几何学的有很多物理学家,其中可以举出例如赫尔姆霍尔茨和马克斯维尔. 这些研究还在继续着;它们不仅有理论价值,而且有实用价值. 对于解决区别信号颜色的问题、纺织工业里的染色问题等等,这些研究给出了精确的数学原理.

2. 我们来讨论在上一节里已经谈到过的另一个例子.

设我们研究某个物理化学体系,譬如气体的混合物、合金等等. 设这个体系的状况由 n 个量决定(例如气体混合物的状况由压力、温度和组成它的气体的浓度决定). 那末就说这体系有 n 个自由度,这表示它的状况可以随着每一个决定这种状况的量的改

变而向 n 个独立的方向变动. 这些决定体系的状况的量就起着它的坐标的作用. 因而它的所有状况的集合可以看作 n 维空间——体系的所谓相空间.

状况的连续改变,即在体系里发生的过程,在这个空间中由曲线表出. 由这种或那种标志分出的状况的各别区域是相空间的区域. 隔开两个区域的状况组成这空间中的曲面.

在物理化学里特别重要的是体系的相空间的那样一些区域的形状和相互邻接,这些区域定性地对应于不同的状况. 隔开这些区域的曲面,对应于这种定性过程,例如溶解、蒸发、沉淀物的沉淀等等. 具有两个自由度的体系的状况由平面上的点表出. 可以取均匀物质作为例子, 这种均匀物质的状况由压力 p 和温度 T 决定; 它们就是表示状况的点的坐标. 于是问题就化成划分区域的曲线,这些区域定性地对应于不同的状态. 譬如在水的情形, 这种区域就是冰、液体水和蒸气的区域(图30). 隔开它们的曲线对应于溶解(凝固),蒸发(凝结),冰的升华(蒸气中冰晶的降落).

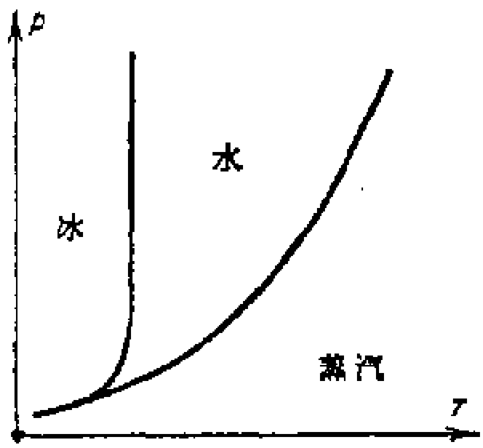


图 30

在研究多个自由度的体系时要用到多维几何的方法.

相空间的概念不仅在物理化学里应用,还在力学体系里应用,一般说来它可以在任何体系里应用, 只要这种体系的可能的状况组成一个连续集合. 在气体运动的理论里,可以讨论例如物质微粒——气体分子的体系的相空间. 一个微粒的运动状况在每一时刻都由它的位置和速度决定, 这就给出了六个量: 三个坐标和速度的三个分量(沿三个坐标轴分解的). N 个微粒的状况由 $6N$ 个量给定, 而且因为分子非常多, 所以 $6N$ 是非常大的数. 这并不使物理学家在说到分子体系的 $6N$ 维相空间时感到任何微小的混乱.

在这个空间里的点表示具有各自的坐标和速度的分子全部的

状况.点的移动表示状况的改变.这种抽象的观点在很深入的理论推导里显得是非常便利的. 总之,相空间的概念已经牢牢地进入严密的自然科学的武器库而且应用在各种各样的问题里.

3. 以上所引的例子已经足以导向如何推广几何对象的结论.

设我们要研究这种或那种对象、现象或者状况的某个连续集合,例如所有颜色的集合或者分子组的状况的集合. 在这种集合里存在的关系可以显得与普通的空间关系相类,例如颜色之间的“距离”或者相空间的区域的“相互位置”. 在这种情况下,抛弃了被研究的对象的定量的特性而且只考虑它们之间的这些关系,我们可以把已知集合看作一种空间. 这个“空间”的“点”就是这种对象、现象或者状况. 在这种空间里的“图形”就是它的点的任何集合,例如虹霓颜色的“曲线”或者在水的状况“空间”里蒸气的“区域”. 这种空间的“几何”就定义为在已知对象、现象或者状况之间存在的那些空间相类的关系. 譬如说,颜色空间的“几何”就由颜色混合的规律和颜色之间的距离决定.

这种观点的现实意义在于,它使我们能够在研究各种各样的现象时利用抽象几何的概念和方法. 因此,几何概念和方法的应用范围是大大地推广了. 空间概念推广的结果,“空间”这个术语在科学里有了两重意义:一方面,这是普通的现实空间——物质存在的普遍形式;另一方面,这是“抽象空间”——某一类对象(现象、状况等等)的集合,在其中存在着空间相类的关系.

应该指出,如果我们对空间观念作若干简化,普通的空间也可以理解成某种状况的集合. 那就是说,它可以理解成极其微小的物体——“质点”的所有可能的位置的集合. 这个附注并非要求给出空间的定义,但是却以更清楚地指出两种空间概念的关系为目的. 抽象空间的概念我们在下面还要说明,而关于抽象几何对普通的现实空间的关系则将在本章最后一节里谈到.

4. 抽象空间的概念在数学本身之中有着最广泛的应用. 在几何学里大家讨论着这种或那种图形的“空间”,例如我们已经提到过的“球体空间”、“直线空间”等等.

特别地，这个方法显得在多面体理论里格外有用．举例说，在第七章§5(第二卷)里曾经提到过关于具有已知展开图的凸多面体存在的定理．这个定理的证明就基于两个"空间"的讨论："多面体空间"和"展开图空间"．把具有已知顶点数的凸多面体的集合看做一种空间，其中的点表示多面体；可能的展开图的集合对应地也解释成一个空间，其中的点表示展开图．用展开图粘成多面体就可确立多面体和展开图之间的对应，即在"多面体空间"和"展开图空间"的点之间的对应．问题就成为证明每个展开图都有多面体与它对应，即一个空间的每个点都有另一个空间的点与它对应．这正好可以用拓扑学的方法来证明．

同样地可以证明一系列别的关于多面体的定理，并且这个"抽象空间的方法"在一系列的情形里(例如在具有已知展开图的多面体存在的定理里)显得是在这种定理的已知证明方法中最简单的．但是可惜这个方法还十分复杂，而且我们在这里也不能给出关于它的更精确的观念．

广义的空间概念还在分析、代数和数论里有着广泛的应用．这首先是函数用曲线的普通表示．一个变数 x 的值通常对应于直线上的点．同样地，两个变数的值对应于平面上的点，n 个变数的值对应于 n 维空间的点；变数的一串值 $x_1, x_2, \cdots, x_n$ 表示具有坐标 $x_1, x_2, \cdots, x_n$ 的点．大家说到这些变数的函数 $f(x_1, x_2, \cdots, x_n)$ 的"变数改变的区域"或者"定义区域"；说到函数的间断点、曲线或曲面，等等．这种几何语言经常被使用着，而且这不仅仅是表示的方法；几何表示使分析的许多事实有了类似于普通空间的"直观性"，而且使得有可能应用推广到 n 维空间的几何证明方法．

在代数里谈到 n 个未知数的方程或者 n 个变数的代数函数时有着同样的情况．在上一段里指出过，n 个未知数的线性函数在 n 维空间里决定平面，m 个这种方程决定 m 个平面，而它们的每一个解表示所有这些平面的一个公共点．若干个平面可以完全不相交，可以相交于一个点，可以相交于一条直线，可以相交于一个二维平面以至一般的 k 维平面．关于线性方程组是否有解的问题被表示

为平面是否相交的问题. 这种几何的处理有一系列的优越性. 一般地说,包含着关于线性方程和线性变换的学问的"线性代数",通常在很大程度上是几何地叙述的,在第十六章里就是这样做的.

5. 在所讨论的一切例子里都提到,这种或那种对象的连续集合被解释成某种空间. 这些对象可以是颜色、这种或那种体系的状况、图形、变数值的全体. 在所有的情形里,对象都由有限个条件给定,因而对应的空间有对应于等于这些条件数的有限的维数.

然而在本世纪初,在数学里开始讨论"无穷维的空间",即这样一种对象的集合,其中每一个对象都不能用有限个条件给定. 这首先是"泛函空间".

把函数的集合解释成某种空间的想法是分析学的新分支——泛函分析的基本思想之一,而且在解决很多问题时显得是非常有效的. 读者将会在专门讲述泛函分析的第十九章里得到这个观念.

可以讨论一个或几个变数的连续函数. 也可以把各种间断函数的类看做"空间",用与从属于问题的解的特征有关的某种方法定义函数之间的距离. 总之,可能的"泛函空间"的个数是无限制的,而且在数学里确实研究着很多这种空间.

完全同样地可以讨论"曲线空间"、"凸体空间"、"已知力学体系的所有运动的空间"等等. 在第七章§5(第二卷)里曾经提到过这样一些定理,在每一个闭凸曲面上至少存在三条闭测地线和每两个点都可以用无穷多条测地线连结它们. 在证明这些定理时利用到曲面上的曲线空间:在第一个定理里是闭曲线的空间,在第二个定理里是连结两个已知点的曲线的空间. 在连结两个已知点的所有曲线的集合里引进它的一种距离,用这种方法就使这个集合变成了空间. 定理的证明基于拓扑学的一些深入的结果对这空间的应用.

现在让我们来表述一般的结论.

在数学里所谓"空间"一般地是指某种对象(现象、状况、函数、

图形、变数的值等等)的任何集合，在这种对象之间有着类似于普通空间关系的关系(连续性、距离等等). 这时把对象的已知集合看作空间，除掉那些决定所考虑的空间相类关系的性质以外，抛弃了这些对象的所有性质. 这些关系决定那种可以称为空间的结构或者"几何"的东西. 对象本身占有这种空间的"点"的地位；"图形"是这空间的点的集合.

组成已知抽象空间的几何学的对象的是空间和其中的图形的那样一些性质，它们由所考虑的空间相类的关系决定. 例如在讨论连续函数的空间时就完全不必研讨个别函数本身的性质. 函数在这里只不过是点，因此"没有部分"，在这意义下除它与别的点的关系外没有任何结构，也没有任何性质；确切地说，这一切都被抛弃了，在泛函空间决定函数的性质的只是它们彼此间的关系——它们的距离和可以从距离导出的其他的关系.

对应于可能的对象集合的多样性和它们之间的各种关系的是在数学里研究的无限多的各种各样的空间. 空间可以按照空间相类关系的类型来分类，这种关系是作为它们的定义基础的. 我们不预备罗列抽象空间的一切不同样的类型，首先来提出两种最重要的类型：拓扑空间和度量空间.

6. 拓扑空间(参看第十八章) 是指任何元素——点的任何集合，在其中定义了一个点对于点集合的接触关系和对应地两个集合(图形)彼此接触或邻接的关系. 这是在普通空间里图形的直观上明显的接触或邻接关系的推广.

罗巴切夫斯基就曾以杰出的智慧指出过，在图形的所有关系之中，最基本的是接触关系. "接触形成立体的特殊的属性，而且我们只保留立体的这个性质，而在论证中不考虑所有其他的性质(不管是实有的或例外的)，那时我们就把立体叫做几何体"[1]. 举例说，圆周上的任意点邻接于圆的所有内点的集合；连通的整个立体的两部分彼此邻接. 拓扑学的进一步发展指明，接触性质正是所

1) 罗巴切夫斯基选集(Н. И. Лобачевский, Собрание сочинений)，第二卷，1949年版，第168页.

有其余的拓扑性质的基础.

接触概念表示了关于点对集合的无限接近的观念. 因而任何对象的每一个集合,只要在其中有着关于连续性、关于无限接近的自然概念,就是一个拓扑空间.

拓扑空间的概念是非常普遍的,而关于这种空间的学问——抽象拓扑学——不是别的,正是关于连续性的最普遍的数学分支.

一般的拓扑空间的严格的数学定义可以用下列方式给出.

一般的拓扑空间是指任何元素——"点"的集合 R,假如在这集合中对于包含在其中的每个集合 M 都定义了接触点,使得下列条件——公理成立:

1) 集合 M 的每个点都算做它的接触点. (认为一个集合与它的每个点接触是十分自然的.)

2) 如果集合 M_1 包含着集合 M_2, 则 M_1 的接触点的集合显然包含 M_2 的所有接触点. (简单地,但是不大精确地说,较大的集合不会有较少的接触点.)

在用这种方法定义某种类型的拓扑空间时,通常还要对于这两个公理添加其他的公理.

利用接触概念容易定义一系列重要的拓扑概念. 这些概念也都是几何学的最基本和最普遍的概念,它们的定义显得十分直观明白. 我们来举一些例子.

1) **集合的邻接** 我们说集合 M_1 和 M_2 彼此邻接,假如其中一个集合至少包含另一个集合的一个接触点. (在这个意义下,例如圆周就邻接于圆的内部.)

2) **图形的连续性或者(像数学家们所说的)连通性** 图形(即点集合) M 叫做连通的,假如它不能分裂成两个彼此不邻接的部分. (举例说,线段是连通的,而去掉中点的线段是不连通的.)

3) **边界** 集合 M 在空间 R 里的边界是指所有这种点的集合,这些点既邻接于 M 本身,又邻接于 M 的补集 $R-M$,即空间 R 的其余的部分. (这显然是十分自然的边界的概念.)

4) **内点** 集合 M 的点叫做内点,假如它不在它的边界上,即如果它不邻接于 M 的补集 $R-M$.

5）连续映射或连续变换 集合 M 的变换叫做连续的，假如它不破坏接触性.（大概不能给出连续变换的更自然的定义了.）

除此之外还可以列上其他重要的定义，例如图形序列收敛于已知图形的概念或关于空间的维数的概念等等的定义.

我们看到，通过接触性可以定义最基本的几何概念. 特别地，拓扑学的意义在于：它给出这些概念的严格的普遍定义，给出严格地应用与直觉地清楚的连续性有关的想法的基础.

拓扑学是关于空间、其中的图形和它们的变换的那样一些性质的学问，它们是用接触关系来定义的.

这个关系的普遍性和基本性使拓扑学成为最一般的几何理论，它到处地深入到数学的各个领域，只要那里提到连续性. 而正是由于它的普遍性，拓扑学在其最抽象的部分已经在实质上超出了几何学的范围. 不过它的基础终究还是现实空间的性质的推广，而且它的最有用的和有力的结果联系着从直观的几何观念里来的方法的应用. 举例说，用多面体逼近一般图形的方法，就由 П. С. 亚历山大洛夫发展了，而且被他推广（虽然是以抽象的形式）到拓扑空间的非常一般的类型.

目前不论是研究什么对象的任何专家都发现了，只要对于他说可以用十分自然的方式引入邻近、邻接的概念，立即就在手边得到现成的、细致分支的拓扑工具，用来做出甚至在它们的特殊现象里也远非显然的结论.

7. 度量空间 是指任意元素——点的集合，在这种点之间定义了距离，即每对点 X, Y 都有与它们对应的数 $r(X, Y)$，使得下列条件——度量空间的公理成立：

1）$r(X, Y)=0$，当且仅当点 X, Y 重合时.

2）对于任何三个点 X, Y, Z，

$$r(X, Y)+r(Y, Z)\geqslant r(Z, X).$$

这个条件叫做"三角形不等式"，因为它完全类似于在欧几里得空间的点 A_1, A_2, A_3 之间的普通距离的熟知性质（图 31）：

$$r(A_1, A_2)+r(A_2, A_3)\geqslant r(A_3, A_1).$$

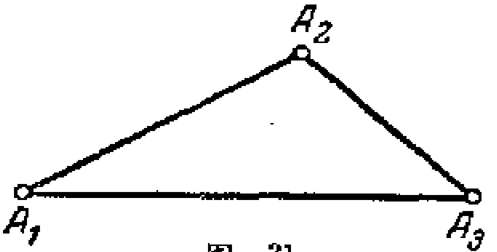


图 31

可以举出度量空间的下列例子：

1) 任何维数 n 的欧几里得空间；

2) 罗巴切夫斯基空间；

3) 具有内蕴度量的任何曲面(第二卷第七章 § 4)；

4) 连续函数的空间 C，其中的距离按下列公式定义：

$$r(f_1, f_2)=\max|f_1(x)-f_2(x)|;$$

5) 将要在第十九章里描述的希尔伯特空间，它不是别的，正是"无穷维的欧几里得"空间.

希尔伯特空间是在泛函分析里应用的最重要的空间；它密切地联系着富里埃级数的理论，而且一般地联系着函数展开成正交函数的级数的理论(在这里坐标 x_1, x_2, x_3, ⋯ 显得是这种级数的系数). 这个空间在数理物理里占有重要的地位，而且在量子力学里具有很大的意义. 原来任何原子体系(例如氢原子)的所有(不仅是静止的)状态的集合，从抽象的观点看来，可以认为是希尔伯特空间.

在数学里讨论的度量空间的例子数还可以大大地加多；顺便提一下，在下一节里我们还要介绍一类度量空间，称为黎曼空间. 但是从以上所引的例子已经足以看出度量空间的一般概念有何等宽广的内容了.

在度量空间里总可以定义全部拓扑概念，而在拓扑概念之外还可以引出别的概念——"度量"概念. 例如曲线长度的概念就是一个. 在任何度量空间里定义长度完全象在普通的情形一样，而且长度的基本性质在这时仍然保留着. 那就是说，曲线的长度是指依次列置在曲线上的点 X_1, X_2, ⋯, X_n 之间距离的和的极限，取极限的条件是这些点在曲线上排得越来越密.

8. 除掉一般的拓扑空间和度量空间以外，在数学里还讨论很多其他类型的空间. 特别说来，在 § 6 里我们已经介绍过一整类的这种空间. 这就是在其中给定了某个变换群的空间(例如射影

空间和仿射空间）．在这种空间里可以定义图形的“相等”．当两个图形中的一个可以用已知群中的变换移到另一个图形上时，就说这两个图形是“相等的”．

我们不预备再进一步定义空间的可能类型了；它们是种类繁多的，而且读者可以去看几何学的各种现代分支的专门著作．

然而，如此地推广几何概念的范围有什么意义呢？为什么必须引入关于连续函数的空间的概念呢？难道不能用普通方法来解决分析的问题，而非要提高到这种抽象的结构不成？

这个问题的一般答案简单地说是这样，在讨论里引入了这种或那种空间，我们就打开了应用能给出很多东西的几何概念和几何方法的途径．

几何概念和几何方法的特点在于，它们在有限的步骤里立足在直观表象之上而且即使以抽象的形式也保持了它们的优点．那些分析学家通过细致的计算才达到的结果，几何学家常常可以立刻就掌握住．作为原始的例子可以来看一下图象，它给出了量之间的这种或那种依赖关系的十分明白的情况．几何方法可以描述成与解析方法不同的概括全部的综合方法．当然，在抽象的几何理论里，直接的直观性是被排除了，但是还保留按照类比的直观想法，还保留几何方法的综合的特征．

读者已经熟悉了几何图的应用，熟悉了复数和复变数函数的几何表示，熟悉了证明代数基本定理的几何论证以及其他的几何概念和几何方法的应用．读者处处可以注意到我们在这里以一般形式说的是什么．我们在§7开端和本节第5段里都曾提出过运用多维几何的方法来证明的定理．我们还要指出从分析中来的一个问题，它已经是在应用泛函空间的概念的基础上来解的了．

在拓扑学里证明了，如果在普通平面上取一个由圆形经过变形而得到的任意区域，然后使它任意地作连续变形，但是必须它整个处在原来的周界之内，则至少有这个区域的一个点在变换后还留在它原来所在的地方．这个事实纯粹是拓扑的．

现在让我们来讨论离几何学十分远的一个问题：考虑一个函

数 $y(x)$，它满足微分方程[1)]

$$y'=f(x, y), \tag{10}$$

而且在 $x=0$ 时取值 $y=0$.

明显地，可以把这个方程换成求下列方程的解：

$$y=\int_0^x f(t, y(t))dt. \tag{11}$$

自然产生这样的问题：一般地说满足这个条件的函数 $y(x)$ 存在吗？

我们换一种方式来看这个问题，我们把每一个连续函数 $y(x)$ 设想成某个抽象空间的点. 计算积分的结果

$$\int_0^x f(t, y(t))dt=z(x)$$

还是 x 的连续函数，即正是我们的抽象空间的"点". 取不同的"点" y，即取不同的函数 $y(x)$，一般说来得到不同的点 z. 因而我们空间的点还是映射成它的点. 方程(11)的解的存在问题化成了这样的问题：在我们的抽象空间里能否找到经过这种变换仍然留在它的原来"位置"的点？

从微分方程理论的自然的问题发现了关于抽象的泛函空间的性质的问题. 与前面提到过的理论的类比告诉我们，谈到的显然是对应空间的拓扑性质.

在这个途径上利用必要的拓扑学的研究，就可以得到关于微分方程解的存在的许多定理是最简短的证明，特别地就能说明方程(10)对于任何连续函数 $f(x, y)$ 确实都有解存在.

§9. 黎曼几何

1. 前面叙述过的、同类现象的连续集合可以解释成某种空间的思想，最初是由黎曼在他的讲演"关于作为几何学基础的假设"里指出的. 这篇讲演是他在1854年在哥丁根大学宣读的. 这是

1) 假定 $f(x, y)$ 是变数 x 和 y 的连续函数.

探索性的讲演，并非为了升到讲师或教授的职位而在数学系里宣读的报告或论文．黎曼在他的讲演里未加推演和数学证明而概括地树立了一种广泛的几何理论的初始观念，这种几何理论现在就叫做黎曼几何．据说，在听众中除掉当时已经年老的高斯外，没有任何人能听懂它．黎曼理论的形式的演算叙述在他的另一个应用于热传导问题的著作中．因而抽象的黎曼几何的产生是与数学物理密切相关的．黎曼的思想是几何学发展中继罗巴切夫斯基以后的极重要的一步．然而黎曼的工作并未立刻得到应有的估价．它的讲演和关于热传导的著作直到他死后的1868年才刊行．应该指出，在1868年出现了由贝尔特拉米给出的罗巴切夫斯基几何的第一个解释，而在1870年出现了由克莱因给的第二个解释．克莱因把关于各种几何（欧几里得几何、罗巴切夫斯基几何、射影几何、仿射几何等等）的普遍见解都表述成了关于在某个群的变换下不变的图形性质的学问．在同一个年代多维几何也在数学中终于得到了巩固．因此，十九世纪的70年代是在几何学历史中的那种突变的时刻，在过去五十年间形成的诸几何观念终于被广大的数学家们所理解而且牢牢地进入了科学领域．于是黎曼的工作有人继续了，而到十九世纪末，黎曼几何已经得到极大的发展而且应用于力学和物理学．到了1915年，爱因斯坦在他的广义相对论里把黎曼几何应用到万有引力的理论，这赋予黎曼几何以特殊的意义，结果使它继续有了急骤的发展和多种多样的应用．

2．有如此光辉的前途的黎曼的思想是十分简单的，假如抛开了数学的探讨而且只注意它的基本的实质的话，这种简单性正是所有伟大的思想所共有的特色．难道从第五条公设的推导来考虑某种可能的几何的罗巴切夫斯基的思想不是简单的吗？难道有机体的演化的思想或者物质的原子结构的思想不是简单的吗？这一切都是简单的同时也是非常复杂的，因为新的思想首先必须为自己开拓广阔的道路，而不是安置在现成的框子里．其次，新思想的多方面的论证，发展和应用需要大量的工作、创造而且不能缺少专门的科学工具．对于黎曼几何这种工具就是它的公式；它们是复杂

的，因而只便于专家们使用．我们不拟讨论复杂的公式，而来注意黎曼思想的本质．上面已经说过，黎曼开始是把现象的任何连续集合看做某种空间的．在这种空间里，点的坐标是在其他现象中间区别对应的现象的量，例如决定颜色 $Ц=xK+y3+zC$ 的强度 x, y, z．如果这种量有 n 个，例如 $x_1, x_2, \cdots, x_n$，则谈到的就是 n 维空间．在这个空间里可以讨论曲线而且引入用微小的（无限小的）步数来测量它们的长度，就象在普通空间里测量曲线的长度一样．

为了用无限小的步数来测量长度，只要给出决定任何已知点到任何与它无限邻近的点的距离的法则．决定（测量）距离的这个法则叫做**度量**．这个法则的最简单的情形是当它与欧几里得空间里的法则相同的情形．这种空间在无穷小范围是欧几里得式的，换句话说，欧几里得几何的几何关系在其中成立，但是只在每一个无限小的区域内；正确地说，它们在任何充分小的区域内成立，但是并非精确地成立，而是区域越小，精确性越大．当空间中的距离按照这种法则测量时，这空间就叫做**黎曼空间**；这种空间的几何叫做**黎曼几何**．因此，黎曼空间是“在无限小范围里”欧几里得式的空间．

黎曼空间的最简单的例子是具有其内蕴几何的任何光滑曲面．曲面的内蕴几何是二维的黎曼几何．实际上，光滑曲面在其每个点邻近与切平面很少差异，而且当我们讨论的曲面区域越小时，这个差异也就越小．所以在曲面的微小区域里的几何与平面上的几何很少差异，而且区域越小，这个差异也越小．然而在较大的区域里，曲面的几何显得与欧几里得几何不同，这在第七章§4（第二卷）里已经指出过，而且也容易从球面和伪球面的例子看出．黎曼几何不是别的，正是曲面的内蕴几何从二维到任何 n 维的推广．类似于只从内蕴几何的观点来讨论的曲面，三维黎曼空间也是在微小区域内是欧几里得式的，而在较大的区域内可以与欧几里得空间不同，譬如说，圆周长度可以不与半径成正比；只是对于微小的圆周这长度才以较好的近似值而与半径成比例．三角形各角的

和不等于两直角(这时在作三角形时起着直线段作用的是最短距离线,即在连结已知点的所有曲线中间有最小长度的曲线).

可以设想现实空间只在与天文尺度相比不大的区域里才是欧几里得式的. 区域越小,欧几里得几何就越精确地成立,但是可以认为(而且它也如此地显得),在很大的尺度里的几何学已经与欧几里得几何有差异. 我们知道,罗巴切夫斯基就已提出了这个思想. 黎曼推广了它,使它不仅对于罗巴切夫斯基几何而且对于任何在无限小范围里是欧几里得式的几何,都有了意义,于是罗巴切夫斯基几何就显得是黎曼几何的特别情形了.

从以上所说的看出,黎曼几何是在由几何学发展所提出的三种思想的综合和推广的基础上产生的. 第一个是关于与欧几里得几何不同的几何可能存在的思想,第二个是关于曲面的内蕴几何的概念,第三个是关于任意维数的空间的概念.

3. 为了说明如何在数学上定义黎曼空间,我们首先回忆一下在欧几里得空间里测量距离的法则.

如果在平面上引进了直角坐标 x, y, 则按照毕达哥拉斯定理,坐标相差 Δx 和 Δy 的两点之间的距离由下列公式表达:

$$s=\sqrt{\Delta x^2+\Delta y^2}.$$

在三维空间里,同理有

$$s=\sqrt{\Delta x^2+\Delta y^2+\Delta z^2}.$$

而在 n 维欧几里得空间里,距离由下列普遍公式决定:

$$s=\sqrt{\Delta x_1^2+\cdots+\Delta x_n^2}.$$

由此容易得出,在黎曼空间里应该如何给定测量距离的法则.这个法则应该与欧几里得的法则重合,但是只在每个点邻近的无限小区域里才是如此. 这导向这个法则的下列表述.

n 维黎曼空间是这样的,在它的任意点 A 邻近可以引进坐标 $x_1, x_2, \cdots, x_n$, 使得从点 A 到无限邻近的点 X 的距离由下列公式表出:

$$ds=\sqrt{dx_1^2+\cdots+dx_n^2}, \tag{12}$$

这里 $dx_1, \cdots, dx_n$ 是点 A 和 X 的坐标的无限小的差. 这还可以

换一种精确的说法，那就是说：从点 A 到任何邻近的点 X 的距离由与欧几里得几何里相同的公式表出，只是具有比点 X 到点 A 的邻近程度为高的精确性，即

$$s(AX)=\sqrt{\Delta x_1^2+\cdots+\Delta x_n^2}+\varepsilon,$$

这里 ε 是比第一项更为微小的量，而且当坐标差 $\Delta x_1, \cdots, \Delta x_n$ 越小时它也越小[1]。

这就是黎曼度量和黎曼空间的数学定义．黎曼度量(即测量距离的法则)与欧几里得度量的差异在于这种法则只在每个已知点邻近成立．此外，这种法则得以简单地表出的坐标，对于不同的点必须取不同的[2]，一般的黎曼度量与欧几里得度量的差异我们以后还要把它弄确切．

黎曼空间在无限小范围内与欧几里得空间重合这个事实，可以用来在其中定义基本的几何量，就类似于在曲面的内蕴几何里用平面逼近曲面的无限小片段时所做的(第二卷第七章§4)．举例说，无穷小体积就可以像在欧几里得空间里一样来表出．有限区域的体积从无穷小体积求和而得出，即从体积的微分求积分而得出．曲线的长度由它的无限邻近的点之间的无穷小距离求和而决定，即由长度的微分 ds 沿曲线求积分而决定．这正是沿曲线使用微小(无穷小)尺度来决定长度的严格的解析表示．从同一个点引出的两条曲线之间的角也可以像的欧几里得空间里一样地来决定．其次，在 n 维黎曼空间里可以给定从 2 到 $n-1$ 的各种维数的曲面．这时容易证明，每个这种曲面同样地也是对应维数的一个黎曼空间，正像普通欧几里得空间里的曲面是二维黎曼空间一样．

还容易证明，黎曼空间总可以表示成充分大维数的欧几里得空间里的曲面，那就是说，对于每一个 n 维黎曼空间，在 $\frac{n(n+1)}{2}$

---

1) 通常公式(12)的精确意义由下列方式表出．设从点 A 引出曲线，因而曲线的点 X 的坐标作为某个变数 t 的函数 $x_1(t), x_2(t), \cdots, x_n(t)$ 而给定．于是这曲线在点 A 处的弧长的微分 ds 由公式(12)表出．

2) 如果可以在整个空间里引进坐标，使得对于任何一对邻近的点，这种测量距离的法则都成立，则空间就是欧几里得空间．

维欧几里得空间里有着 n 维曲面，它从其内蕴几何的观点看来与这个黎曼空间并无区别(至少在它的一定的有限部分里是如此)。

4. 为了在黎曼几何里得出各种几何量的解析表示，首先必须给出在黎曼空间里测量长度的法则的一般表达式，它必须不专门依赖于每个点的独特坐标。可是公式(12)只是在点 A 的特别选取的坐标系里才对这个点成立的，因而当从一个点过渡到另一个点时必须改变坐标系，这当然是不方便的。要消除这个不便原来也还容易，也就是说可以证明下面的事实。

设在黎曼空间的某个区域里引进了任何坐标 $y_1, y_2, \cdots, y_n$。于是从具有坐标 $y_1, y_2, \cdots, y_n$ 的点 A 到具有坐标 $y_1+dy_1, y_2+dy_2, \cdots, y_n+dy_n$ 的点 X 的"无穷小距离"或者说"长度元素" ds 就由下列公式表出：

$$ds=\sqrt{\sum_{i,k=1}^{n} g_{ik}\,dy_i\,dy_k} \text{ 或者 } ds^2=\sum g_{ik}\,dy_i\,dy_k, \qquad (13)$$

这里系数 g_{ik} 是点 A 的坐标 $y_1, y_2, \cdots, y_n$ 的函数。

后一个公式右边的式子叫做坐标微分 $dy_1, \cdots, dy_n$ 的二次形式[1]。它的展开式可以写成

$$\sum g_{ik}\,dy_i\,dy_k=g_{11}\,dy_1^2+g_{12}\,dy_1\,dy_2+g_{21}\,dy_2\,dy_1+g_{22}\,dy_2^2+\cdots.$$

因为 $dy_1\,dy_2=dy_2\,dy_1$，所以不妨把第二和第三两项看做是相同的：$g_{12}=g_{21}$ 而且一般地 $g_{ik}=g_{ki}$；这是可以的，因为重要的只是它们的和 $(g_{ik}+g_{ki})\,dy_i\,dy_k$。

除掉所有微分 dy_i 都等于零的情形，二次形式(13)总是正的，因为显然有 $ds^2>0$。

逆命题也成立。那就是说，如果在引进了坐标 $y_1, y_2, \cdots, y_n$ 的 n 维空间里，长度元素由公式(13)给定，并且公式右边的二次形式是正的(即除掉所有 $dy_i=0$ 的情形以外，总大于零)，则这空间就是黎曼空间。换句话说，在每个点 A 邻近可以引进新的特殊的坐标 $x_1, x_2, \cdots, x_n$，使得在新坐标里这个点处的长度元素由最简

1) 若干个量的二次形式是指这样一种代数式子，它对于这些量说是齐次的二次多项式。

单的公式(12)表出.

$$ds^2=dx_1^2+dx_2^2+\cdots+dx_n^2.$$

因此，黎曼度量(即在无穷小范围内是欧几里得式的长度的定义)可以用具有系数 g_{ik} 的任何正的二次形式(13)给定，这里 g_{ik} 是诸坐标 y_i 的函数. 这就是给定黎曼度量的普遍方法.

在黎曼空间里的曲线是这样给定的，点的所有 n 个坐标都随着在某个区间里改变的一个参数 t 而改变

$$y_1=y_1(t),\ y_2=y_2(t),\ \cdots,\ y_n=y_n(t)\quad (a\leqslant t\leqslant b).\qquad (14)$$

曲线的长度由积分表出:

$$s=\int ds=\int\sqrt{\sum g_{ik}\,dy_i\,dy_k}.$$

在曲线由方程(14)给定的情形，

$$dy_1=y_1'dt,\ \cdots,\ dy_n=y_n'dt,$$

因而

$$s=\int_a^b\sqrt{\sum g_{ik}y_i'y_k'}\,dt.\qquad (15)$$

因为 g_{ik} 是坐标 $y_1, \cdots, y_n$ 的已知函数，而这些坐标又以已知的方式按着公式(14)而依赖于 t，所以公式(15)里在积分记号下的是对于已知曲线完全确定的 t 的函数. 因此，这函数的积分有确定的值，因而曲线本身就有确定的长度.

连结两个已知点 A, B 的最短曲线的长度采用作为这些已知点之间的距离. 这曲线——测地线——本身占有类似于直线段 AB 的地位. 可以证明，在微小区域里任何两个点由唯一的最短线连结. 求测地线(最短线)的问题本身是使积分(15)具有极小值的问题. 这是读者在第八章(第二卷)里已经接触过的变分法的问题. 变分法的正常的应用就可以用来引出决定测地线的微分方程和对于任何黎曼空间确定测地线的一般性质.

让我们来证明前面指出过的基本断言: 黎曼度量在任何坐标里都由普遍公式(13)给定.

设在黎曼空间的某个区域里引进了任何坐标 $y_1, y_2, \cdots, y_n$. 我们在这个区域里取任意点 A，而且设 $x_1, x_2, \cdots, x_n$ 是点 A 的特殊坐标，使得在这些

坐标里点 A 处的长度元素由公式(12)表出，这就是说，

$$ds^2=dx_1^2+dx_2^2+\cdots+dx_n^2. \tag{16}$$

坐标 x_i 由坐标 $y_j(i,\ j=1,\ \cdots,\ n)$ 表出的公式是：

$$x_1=f_1(y_1,\ y_2,\ \cdots,\ y_n),$$
$$x_2=f_2(y_1,\ y_2,\ \cdots,\ y_n),$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$x_n=f_n(y_1,\ y_2,\ \cdots,\ y_n).$$

于是
$$dx_1=\frac{\partial f_1}{\partial y_1}dy_1+\frac{\partial f_2}{\partial y_2}dy_2+\cdots+\frac{\partial f_n}{\partial y_n}dy_n.$$

对于 $dx_2,\ \cdots\cdots,\ dx_n$ 也有类似的式子．把这些式子代入公式(16)，于是在计算了右边的平方而且把具有 $dy_1^2,\ dy_1dy_2,\ dy_2^2$ 等等的各项归并以后，我们就得到下列式子：

$$ds^2=g_{11}dy_1^2+2g_{12}dy_1dy_2+g_{22}dy_2^2+\cdots\cdots+g_{nn}dy_n^2,$$

这里系数 $g_{11},\ g_{12},\ \cdots,\ g_{nn}$ 都是由偏导数 $\dfrac{\partial f_i}{\partial y_j}$ 表出的式子（这些式子的形状对于我们说不是主要的）．但是这不是别的，正是写成展开式的公式(13)，因而我们的断言证明了．

现在让我们来证明，反之，公式(13)决定黎曼度量，即当对于每个点都选取它的特殊坐标 x_i 时，总可以引向简单的形状(16)．设

$$ds^2=\Sigma g_{ik}dy_idy_k,$$

并且 g_{ik} 是坐标 $y_1,\ \cdots,\ y_n$ 的函数而且右边的二次形式是正的，我们取任何点 A 而且只在这个点处讨论已知的二次形式．于是系数 g_{ik} 都是已知数，而二次形式所依赖的变数是 $dy_1,\ \cdots,\ dy_n$．从代数知道，每一个正的二次形式（具有任何的数字系数）都可以用线性变换化成平方和（参看第十七章）[1]，即存在这样的变换

$$dy_1=a_{11}dx_1+\cdots+a_{1n}dx_n,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$dy_n=a_{n1}dx_1+\cdots+a_{nn}dx_n,$$

当把这些式子代入公式(13)以后，我们得到

$$ds^2=dx_1^2+\cdots+dx_n^2.$$

如果我们按照下列公式把坐标 $y_1\cdots,\ y_n$ 换成 $x_1,\ \cdots,\ x_n$：

$$y_1=a_{11}x_1+\cdots+a_{1n}y_n,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

1) 我们这里的变数具有微分的形状是不要紧的，我们完全可以把它们看作某些独立的变数．

$$y_n = a_{n1}x_1 + \cdots + a_{nn}x_n,$$

则微分 dy_i 也正好由微分 dx_i 按照公式(17)表示. 因此, 这种坐标变换解决了所提出的问题: 在坐标 $x_1, \cdots, x_n$ 里, 在我们所取的点处的微分的平方 ds^2 由简单的"欧几里得式的"形式(16)表出. 因而也就证明了, 一般的公式(13)确实给定黎曼度量.

5. 欧几里得空间是黎曼空间的最简单的特殊情形[1]. 黎曼几何的主要任务是给出区别一般的黎曼空间与欧几里得空间的解析表示, 定义所谓非欧几里得的黎曼空间的测度. 这个测度就是空间的所谓曲率.

首先应该指出, 关于空间的曲率的概念, 完全与这空间在那样一个高维的包容它的空间里如何弯曲的观念无关. 曲率可以在已知空间本身之内定义, 而且可以在它的内蕴几何性质的意义下表出它与欧几里得空间的差异. 这必须理解清楚, 不要把关于空间的曲率的概念与它的任何构成法连在一起来看. 当说到我们的现实空间有曲率时, 这只是表明它的几何性质与欧几里得空间的性质有区别. 但是这完全不表示我们的空间处在某个更高维的空间里, 而它就在那里弯曲着. 这种观念与黎曼几何对现实空间的应用毫无关系, 而属于空洞的幻想的范围.

黎曼空间的曲率的概念推广了曲面的高斯曲率的概念到 n 维. 在第七章 § 4(第二卷)里曾经说过, 高斯曲率表示曲面的内蕴几何对平面上的几何的差异程度, 而且可以从纯粹内蕴几何的观点来讨论. 它不是别的, 正是代表已知曲面的二维黎曼空间的曲率.

作为例子, 我们来回忆出现高斯曲率的两个内蕴几何的公式. 设在曲面上某个点 O 邻近有一个微小的三角形, 它的边是测地线; 设它的角是 α, β, γ. 面积是 σ. 量 $\alpha+\beta+\gamma-\pi$ 表示它的各角

1) 在欧几里得空间里, 长度元素在直角坐标里是由公式(16)表出的: $ds^2=\Sigma dx_i^2$. 如果过渡到其他的坐标, 则根据第4段的结论, ds^2 由某个二次形式(13)表出. 因此, 在欧几里得空间里, 长度元素在任意坐标里的表达式也是一般的公式(13). 然而, 欧几里得空间与任何其他空间有区别, 因为在欧几里得空间里可以引进这样的坐标(这就是直角坐标), 使得公式(16)在同一些坐标里到处成立, 而不像在一般的黎曼空间里, 这只是在这个或那个点的邻近才成立.

之和与平面上的三角形的各角之和的差别.

如果三角形向点 O 缩小，则 $\alpha+\beta+\gamma-\pi$ 对它的面积 σ 的比值趋向点 O 处的高斯曲率 K. 换句话说，对于微小三角形有

$$\frac{\alpha+\beta+\gamma-\pi}{\sigma}=K+\varepsilon,$$

这里当三角形向点 O 缩小时，ε 趋于零. 这正好表明，高斯曲率是曲面上的三角形的各角之和与平面上的三角形的各角之和的差异程度.

我们再来讨论曲面上有中心在点 O 处的圆（即在曲面上的距离的意义下与点 O 等距离的点的轨迹）. 如果 r 是圆的半径，l 是圆周长度，则

$$l=2\pi r-\frac{\pi}{3}Kr^3+\varepsilon,$$

这里 K 还是点 O 处的高斯曲率，ε 表示与 r^3 相比是微小的量.

这里高斯曲率表出了微小圆周的长度与 $2\pi r$ 的差异程度，而这两者在欧几里得几何里是相等的.

黎曼空间的曲率起着类似的作用. 例如它可以用下列方式定义. 在已知的黎曼空间里有着一个光滑曲面 F，它由通过已知点 O 的测地线组成. 这个曲面的高斯曲率就采用作为空间在点 O 处向着曲面 F 的方向的曲率. 一般地说，这个曲率不仅在不同的点 O 处是不同的，而且对于通过同一个点 O 的不同的测地线曲面 F 也是不同的. 因此，空间在已知点处的曲率并不由一个数决定. 黎曼就已经引出过联系在同一个点处的不同曲面 F 的曲率的一般规律. 由于这个原因，空间在一点处的曲率由一组数——所谓曲率张量决定.

但是我们没有可能在这里说明这个原因，因为这需要引用大量的数学工具. 重要的是要理解曲率是黎曼空间的非欧几里得性的尺度；它在这空间内部决定它的度量与欧几里得空间的度量的差异程度. 例如它决定三角形各角之和与 π 的差异和圆周长度与 $2\pi r$ 的差异. 一般地说，它在不同的点处有不同的值，而在同一个

点处它并非由一个数而是由某一组数给定.

黎曼空间就其本身的性质说可以不是均匀的，那时在这空间里就不能自由地移动图形而不改变它的各点之间的距离. 因而就发生这样的问题：在何种黎曼空间里可以自由地移动图形并且具有与在欧几里得空间里相同的自由程度. 这种空间是最均匀的黎曼空间.

原来欧几里得空间是均匀的无曲率的空间(零曲率的空间). 另一类型的均匀空间是罗巴切夫斯基空间. 因此罗巴切夫斯基空间也像欧几里得空间一样是一般的黎曼空间的特别情形.

一般地说，可以在其中自由地移动图形的空间都是常数曲率的空间. 在这种空间里，曲率在所有的点处对于所有的测地线曲面有同一个值. (代替随点而变的"曲率张量"，这时曲率由对于所有的点都相同的一个数给定.)零曲率的空间是欧几里得空间，负曲率的空间是罗巴切夫斯基空间，正曲率的空间与 $n+1$ 维欧几里得空间里的 n 维球面有相同的几何.

6. 黎曼几何并不需要长久地等待应用，我们已经说过，黎曼自己就把它应用于解决热传导问题的形式的计算；但是这只是应用它的公式，而不是应用在无穷小区域里具有欧几里得式的距离测量的抽象空间的观念. 这种应用首先是对于颜色空间而给出的，在颜色空间里利用黎曼度量，邻近颜色之间的距离就能表示了；颜色空间就能解释成为独特的三维黎曼空间.

黎曼几何在力学里也有重要的应用. 为了说明它的实质，我们先来讨论点沿曲面的运动. 设想有个质点，例如小弹子，它可以自由地沿着光滑曲面移动，但是不能离开曲面. 点好象是在曲面本身之内运动. 在曲面上可以引进某种坐标 x_1, x_2；于是点的运动就由它的这些坐标对时间的依赖关系完全决定，而速度则由坐标的改变速度，即由它们对时间的导数 $\dot{x}_1$, $\dot{x}_2$ 来决定. 结果点好象在二维空间里运动，但是这空间不是欧几里得空间，而是有其自己的几何——曲面的内蕴几何的空间. 运动的规律可以变换成这样，使得在其中只出现点在曲面上的坐标 x_1, x_2 以及它们的一

阶和二阶导数.

如果有力作用在点上，则它的垂直于曲面的分力与曲面的反作用力抵消，因而只留下与曲面相切的分力；它已经是只沿曲面而作用了. 因此，作用于点上的力可以认为是在曲面本身之内作用的. 曲面的内蕴几何是黎曼几何的特殊情形. 因而点沿曲面的运动是在二维黎曼空间里的运动. 它的运动规律有与普通的运动规律相同的特性，不同的只是在其中考虑的是曲面的内蕴几何. 从第七章§4(第二卷)里提到过的下列事实，这变得格外清楚：在曲面上按照惯性而且没有摩擦力而运动的点，以定常速度沿着测地线而运动. 因为测地线在曲面上起着直线的作用，所以所说的事实类似于惯性定律；这是同一个惯性定律，但是是关于曲面上的运动的，或者抽象地说是关于二维黎曼空间里的运动的.

当然，在这种抽象的观念里暂时还看不到任何优越之处，因为谈的只是沿普通曲面的运动.

当我们转到由两个以上的量给定其位置的力学体系时，这种抽象观点的效益立刻显现了. 那时用点在曲面上的运动来表示它们的运动变得不可能了. 在§7里谈到图表方法如何在多维空间的抽象观念中发展时，我们已经遇到过这种情况了.

总之，设给了一个力学体系，它的位形，即其各部分的位置，由 n 个量 $x_1, x_2, \cdots, x_n$ 给定. 如果谈的是若干个质点的体系，则它们的位置就由给定它们的所有坐标（每个点三个坐标）来决定. 另一个例子是陀螺(一个轮子绕轴旋转，这轴本身又可以绕不动的点而转动). 陀螺绕轴的旋转由旋转角给定，而轴的倾斜又由两个角给定，这两个角是由轴与两个已知方向组成的. 得到的一共是决定这种陀螺的位置的三个量（图 32).

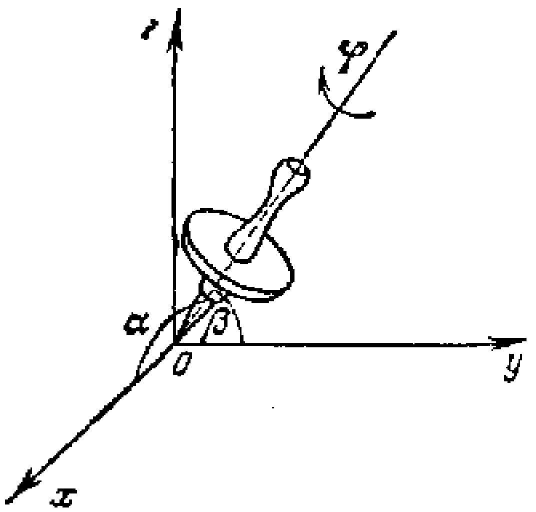


图 32

体系的每个位形——体系各部分的位置——可以理解成它的所有可能的位形的空间中的"点". 这就是所谓体系的位形空间[1]. 它的维数等于决定位形的量 $x_1, x_2, \cdots, x_n$ 的个数. 这些量是点在位形空间里的"点"的坐标. 对于譬如说由三个质点组成的体系, 我们得到三个点的各三个坐标——一共是九个坐标. 对于陀螺的情形我们有三个坐标——三个角, 因而陀螺的位形空间是三维的.

体系的运动由点在位形空间里的运动表示. 这个运动的速度由坐标 $x_1, x_2, \cdots, x_n$ 的改变速度决定.

关于这种空间, 它们的拓扑结构还将在第十八章里谈到. 而这里我们想要强调的是, 在位形空间里可以引进测量距离的特殊法则, 它是与体系的力学性质密切有关的. 那就是说, 如果体系的动能由下列公式表出:

$$T=\frac{1}{2}\sum_{i,k=1}^{n} a_{ik}\dot{x}_i\dot{x}_k,$$

这里 $\dot{x}_i$ 是对应的坐标的改变速度, 则无穷小距离的平方就由下列公式给定:

$$ds^2=\sum_{i,k=1}^{n} a_{ik}\,dx_i\,dx_k.$$

因此, 位形空间变得是黎曼空间. 这时不仅体系的运动由点在位形空间里的运动表示, 而且描述体系运动的方程本身也与这个点的运动方程重合; 总之, 体系的力学作为点在位形空间里的力学而表出. 特别地, 体系按着惯性即无反作用力的运动(类似于陀螺的自由旋转), 就由点在这个空间里沿着测地线的等速运动表示.

这个表示在许多情形里显得是方便的, 而且它和它的某种推广和变形在理论力学里都要用到.

因此, 黎曼几何可以用作物理现象的抽象几何的描述方法. 这种描述完全不是任意的, 也不是数学的智力游戏, 它反映了所讨论的现象的现实的规律性, 只是以抽象的形式来反映它们罢了. 这就是物理现象的每一个数学描述的本质. 差别只在于这里应用的是

---

1) 不要把它与 §8 第 2 段里提出的"相空间"相混. 在相空间里, 点不仅由位置决定, 还由体系的点在每一时刻的运动速度决定.

更有力、更细致的抽象，但是实质还是一样的.

黎曼几何在相对论里得到最出色的应用. 但是关于这一点我们将在下一节里说到，因为在这时谈的是抽象空间对现实空间性质的关系的重要而又困难的问题.

7. 最近 30 年来，各种非欧几里得空间的几何学在各个方向得到重大的发展和推广，产生了包括黎曼几何作为特别情形的一些新的理论. 其中第一个是所谓芬斯勒几何，这种几何的观念在黎曼以前就已出现[1]；然后是伟大的法国几何学家卡当的连结了黎曼几何与克莱因的爱耳浪根纲领的空间的一般理论，以及其他的理论. 因为不可能谈到几何学的这些新方向，我们只预备指出一下，它们主要是借助于专门适用于它们的解析工具而展开的. 很多苏联数学家参加了这些新方法的发展工作；其中有由拉舍夫斯基创立的所谓新的“半度量”几何，瓦格涅从空间曲线理论的最一般的问题扩展到非欧几里得几何在力学里的应用的研究，等等.

§10. 抽象几何和现实空间

1. 随着前面叙述的从罗巴切夫斯基开始的几何观念的发展，我们深入到抽象的空间之内，而且远离最初的几何对象——容纳一切现象的那个现实空间. 现在让我们回到普通意义下的这个空间，而且我们的任务是要说明，究竟是什么东西为了认识这个空间的性质而使抽象几何有了发展.

我们知道，几何学是从经验、从物体的空间的形式和关系的经验研讨，即从土地、器具容积等等的测量而产生的. 因此，从来源说它像力学一样是一种物理的理论. 欧几里得几何的公理是从长时期的经验得来的清楚地表述出来的结论；它们表出了自然的规律，因而它们可以叫做几何学的规律，正像力学的基本规律现在也

1) 芬斯勒是德国数学家，他在 1916 年开始上述几何的详细的研究.

常常叫做力学的公理一样[1]．但是决不能断言，这些规律是绝对地精确的，而且完全不需要随着新的经验资料而更加以精确化和加以概括；空间的现实性质可以在某种程度上与欧几里得几何所给出的结论有差异．

这些论点我们早已提出过了，现在它们大概是十分明显了罢．但是在一百年以前，当罗巴切夫斯基的思想还没有获得普遍的承认时，却并不是这样．在罗巴切夫斯基和高斯以前，任何人也没有想到欧几里得几何可以是不完全精确的，而空间的现实性质可以是另一些样子的．罗巴切夫斯基是把它的几何当做现实空间的可能性质的理论来发展的．后来黎曼和其他一些学者也提出了关于空间的可能性质的问题，关于在精确测量时可能发生的测量长度的可能法则的问题．一般地说，抽象几何在其某些部分可以看作空间的可能性质的理论．然而假如不是爱因斯坦 1915 年在他的广义相对论里证实了罗巴切夫斯基和黎曼的观念，这一切还会留在假设的范围里．相对论断言，现实空间的几何确实与欧几里得几何有某些差异，而且这正可以象罗巴切夫斯基所希望的那样以天文尺度发现．

在刚才关于空间所说的话里，至少隐藏了三个难点．为了说明这些难点，归根到底还是要引向抽象几何对物理几何（即现实空间的性质）的关系的问题．

第一个难点在于，设想现实空间的性质如何地和在什么意义下与欧几里得几何所给的结论不同．我们过分习惯于欧几里得几何了，以致为了容易想像一些什么别的东西，当然必须加以解释了．

第二个难点包括在"现实空间的性质"这句话本身之中．空间

1）抛弃了公理的最初内容的公理的抽象理解，是在 50 年前才发生的，它完全没有改变欧几里得几何公理反映自然规律这个事实，在说到公理而不是说到几何学或力学规律时，人们是希望把这些科学的逻辑的演绎结构提到第一位，但是这些科学并不因此而失去其经验基础．一个理论的某个命题叫做公理，假如它被采用作为这理论的演绎结构的基础．于是这理论的其他的命题（定理）就可以用逻辑推理的方法从基本的命题（公理）导出．

被孤立地理解成空的和均匀的．表面上看来在空间概念本身之中已经包括了关于它的均匀性的观念，而且空的空间即“真空”又能有些什么性质呢？我们说“空间的性质”时并未考虑到这些问题，但是只要略加考虑，上述难点就可以变得十分明显了．

第三个难点包括在这种或那种几何的真实性的概念之中．表面上看来事情是很简单的：合乎实际情形的几何才是真实的．这当然是对的．但是在另一方面我们却看到，譬如说圆内的几何可以理解成罗巴切夫斯基几何，因为圆内的每一个几何事实都可以叙述成罗巴切夫斯基几何的定理．因此，同一个几何事实既可以叙述成欧几里得几何的定理，也可以叙述成罗巴切夫斯基几何的定理．这说明两种几何都合乎实际情形．那末在什么意义下它们之中的那样一个是真实的呢？而且为什么我们仍然认为在圆内事实上成立的是欧几里得几何，而罗巴切夫斯基几何只是在其中有了表示、得到了解释呢？

很明白，这些问题确实有一定的困难，才会使某些伟大的数学家在当时被难住．

解说必须从所说难点中的第二个开始，因为只有理解了什么是“空间的性质”，才能来解决其他的难点．

2．几何学的对象——“空间的性质”——由现实物体的性质、它们的物质关系和形式组成．在现实空间里，“位置”、“点”、“方向”等等都由物体决定．“这里”和“那里”、“向这里”和“向那里”等等也只有在与这个或那个物质对象的联系和关系中才有意义．“这里”可以是指“在地上”、“在这个房间里”等等的意义；总之，“这里”总是指由这些或那些物质特征决定的位置．同样地，譬如说直线并不实际存在，而存在的只是拉紧的线，或者尺的边沿，或者光线．而一般的直线，“直线本身”是反映了这些物质直线的公共的性质的抽象，正象我们说“一般的房屋”是反映了房屋的公共性质的抽象一样；“一般的房屋”并不能在各个现实的房屋之外而独立存在．

空间性质的这个客观的特点由辩证唯物主义的熟知命题表出．空间是物质存在的形式．对象的形式由其各部分的联系和关

系决定．空间的结构是一系列物体的关系和现象的规律性．这就是空间关系、对象的空间顺序、它们的相互关系等等．但是正像每一个形式都不能脱离内容一样，即使是在抽象方面和在一定的范围内，空间也不能脱离物质．关于"孤立的"空间、没有物质的空间的观念是不能滥用的抽象．现实的空间关系和形式："这里"、"在…之间"、"在…之内"、"直线"、"球"等等，总是物体的关系和形式．而几何学是抽象地来讨论它们的．从具体性得出的这种抽象是必要的，因为否则就不能认识在对象的各种具体关系中所共有的东西．但是不能把这种抽象绝对化，不能用抽象的概念来代替客观的现实性．

在完全没有物质的踪迹的绝对空的空间里，无法区别任何位置、任何方向，因此也就没有位置，没有方向．因而绝对空的空间变成什么也没有了．即使在空的空间的抽象观念里，也隐含着在其中能区别不同的位置和方向的意思．换句话说，在空间的抽象观念里就包含着位置、方向、距离的可区别性，这种可区别性在现实空间中存在就是由于这空间是不可分割地与物体相联系的．

总之，空间是物质存在的形式；因此，"空间的性质"是物质的性质，是物体的已知关系、相互位置、大小等等的性质．

其次，为了使得几何定理具有物理意义，必须知道在其中应该如何理解"直线"、"距离"和其他的几何概念．在§4里我们看到，同一个几何定理许可不同的解释．

因此，在比较几何学与经验时，必须尽可能精确地确定几何概念的物理意义，因为几何学描述现实空间的性质是有条件的，这条件就是它的概念具有对应的物理意义．没有这个物理意义，几何学的定理就只有抽象数学的、形式的特性了．这已经包含了上面所说的第二个难点的解答了．这个难点的产生，是因为想用"纯粹"空间、"自在"空间，来代替不可分割地与物质相联系的现实空间，然而前面那种空间只不过是抽象罢了．

8．现在就容易理解另外两个难点如何解决了．

第一，怎么可以设想现实空间的性质与欧几里得空间不同呢？

设我们希望检验欧几里得几何的某个断言，例如三角形各角之和等于180°，或者圆周长度等于$2\pi R$. 为了检验第一个断言，必须确定，讨论的是怎样从物理上决定的三角形，它们的角又如何决定. 设三角形的边是真空中的光线. 在这种情况下，很精确的实验会显示出三角形各角之和与180°的差异，这并不是什么不可思议的事. 同样可以设想，用同一个尺度来测量半径和圆周会得出一种结果，它并不是精确地满足关系$l=2\pi R$的. 比方说这在地球表面上就是这样，在地球表面上圆周长度并不与半径成正比，而是较慢地增长，而且当半径等于子午线的一半时达到其极大值. 对于这一点可能有异议：谁不知道在地球表面上起直线作用的是大圆弧，因而半径在这里有另一个意义，所以我们的结果并不与欧几里得几何矛盾. 然而，根据相对论，在较大质量的物体邻近，圆周长度对"实在的"半径的比值仍然与2π有一些差异，那就是说，均匀的球状物体的赤道对它的半径的比值有下列近似公式：

$$\frac{\text{圆周长度}}{\text{半径}}=2\pi\left(1-\frac{kM}{3Rc^2}\right),$$

这里M是物体的重量(以吨计算)，R是物体的半径(以公里计算)，$c=300,000$公里/秒是光速，k是引力常数，在所取的测量单位下等于66.6×10^{-18}.

这样一来，我们看到，圆周长度对半径的比值不等于2π，而稍稍小些. 从计算知道，在太阳表面上这个比值与2π的差异接近于0.000004，而在平均密度有水的密度的50,000倍大的天狼星伴星的表面上，这差异已经达到0.00014.

当然还可以反驳说，这一切仍然无法设想，在空间的直观表象中总还是欧几里得式的. 这种反驳不应该使我们动摇，首先因为科学的任务并不是为了给出现象的直观表象，而是为了得到理解它们的结果. 直观的表象是有限制的，不能超出使我们的器官获得感觉的惯常形象的范围. 所以我们如果不用一些什么模型来代替现象，就不能直观地设想紫外线、无线电波的传播、电子在原子里的运动以及很多其他的东西. 但是这完全不表明这些现象不能

被我们所理解．相反地，例如无线电技术的成就就明白地指出，我们完全掌握了无线电波，因此也就是十分好地理解了它们．其次，什么可以设想、什么不可以设想这一问题的解决依赖于想象的习惯和训练．能够设想一种头朝下（相对于我们来说）走路的对蹠人吗？现在我们可以设想这个，而在当时，对蹠人的"不可想象"正是反对地球的球形的论据哩．

4．现在我们转到最后一个难点，即哪一个几何才算是真实的问题．在这个问题的提法中就已指出，圆内的几何事实既可以叙述成欧几里得几何的定理，也可以叙述成罗巴切夫斯基几何的定理．因此，这两种几何都合乎实际情形，即它们都是真实的，而且在这里终究并无任何不可思议之处．同一个现象总可以用不同的方式来描述；同一个量可以用不同的单位来测量；同一条曲线可以用依赖于坐标系的选取的不同的方程来给定．完全同样地，从几何关系分出的已知部分（例如在所讨论的例子里的圆内的关系）可以用不同的方法来描述，但是我们提的问题不是关于几何事实的某个孤立的部分，而是关于最普遍的空间关系．空间是物质存在的普遍形式，因此，在提出关于空间性质的问题时，都不能人工地分出任何事实范围．

在几何量的问题的这种提法下，不能认为几何事实能够脱离与它们有必要联系的其他现象而独立．举例说，线段的长度是用移动刚体尺的方法来决定的，因而长度的测量必须与刚体运动相联系，即与力学相联系．几何学不能与力学分离．然而在§4里指出过，在圆内测量长度在罗巴切夫斯基几何的解释下是以完全另外的方式进行的；这时弦有了无穷的长度．显然，这样定义的测量不符合于关于测量的原来的观念，后者却是在被比较的现实物体的力学位移的基础上产生的．一般地说，在欧几里得几何的意义下相等的图形，按照欧几里得几何的来源，就是经过力学的运动可以彼此迭合的图形．在罗巴切夫斯基几何的解释里，相等是另外定义的，在那里占有移动的地位的是另一种变换．因此，如果圆内的几何事实是在它们与力学的必要联系中考虑的，则我们应该

承认，在圆内实际上成立的正是(具有较大精确性的)欧几里得几何.

欧几里得几何是以刚体的普通力学运动作为移动的几何. 这就是为什么人们首先发现欧几里得几何而不是任何其他的几何的道理. 但是物理学的发展现在已经得出这样的结论，牛顿的力学定律连同欧几里得几何的规律都只是更精确和更普遍的规律的近似罢了. 在相对论里实现的几何规律从欧几里得几何过渡到黎曼几何的这种改变，不仅在力学里起作用；这在电磁现象和实验里也有着不更小的(假如不更大的话)意义. 几何学作为与物理学相联系的关于空间性质的科学，它依赖于物理学，而且只有在抽象中、在一定范围内才能离它而独立.

罗巴切夫斯基已经指出了几何学对物理学的依赖性和空间性质对物质的依赖性，它预见了几何学的规律在过渡到新的物理现象的领域里可能改变. 站在这个唯物主义观点的对立面，著名的数学家庞加莱在较近的时期还断言，这种或那种几何学的选择只是由单纯的想象或者象著名的唯心主义者马赫的所谓"思维的经济"所决定的. 在这个基础上，庞加莱甚至预言，要科学放弃光的直线传播定律，也要比放弃欧几里得几何更容易些，因为后者是"太简单了". 然而庞加莱只差三年没有能活到广义相对论的终于建立，在广义相对论里正好得到相反的结果：放弃了欧几里得几何，但是却保留了光的传播定律，当然是在更普遍的形式下，光是沿着测地线而传播的. 这并不是马赫主义者的观点的第一次失败. 在本世纪的前夕，马赫主义者和其他相近流派的唯心主义者断言，原子只是思维的符号或者这一类的什么东西，但是决不会是客观的现实性. 然而没有过去多少年，原子的现实性就被人用最可信的实验证明了. 不管这一切，几何学的唯心主义认识，也正像在其他科学中一样，仍然保留在某些资产阶级学者中间. 对于这些学者说，资产阶级的意识形态还是有力的客观真理.

总之，我们引向下面的结论. 从诸般事实分出的同一部分，一般地说可以作不同的描述，而且所有这种描述都是真实的，只要它

们反映了现实性．然而最具普遍性的几何事实决不能在与其他现象割裂的情况下讨论．空间性质之所以可以确立，就因为它是物质存在的普遍形式．而在与物理学的联系中接受几何学时，我们必须使它们彼此适应，于是各种“几何”之间的实质的差别也就暴露了出来，这种差别在与物理学割裂的时候可能只是由于较多或较少单纯性而有所不同罢了．欧几里得几何的出现，并非因为它比其他几何简单，而是因为它对应于力学．而现在就是由于物理学的发展，在相对论里才转入更复杂的几何——黎曼几何．

一句话，对于现实空间的性质说来，真实的是这样的几何，它充分精确地反映了最具普遍性的空间关系的性质，因此它不仅符合纯几何的事实，而且也符合力学以至于整个物理学．

5．上面关于相对论说得不多，同时没有提到它的主要内容．在理解空间问题时，它比罗巴切夫斯基和黎曼关于这个问题所想到的要走得远得多．

相对论的最主要的和最基本的命题是这样的，空间与时间有不可分割的密切关系，它们共同组成物质存在的统一的形式：时空的四维流形．世界上的事物由位置和时间刻画，因此就要用到四个坐标：三个空间坐标和第四个时间坐标——事物的时间．在这意义下，事物就组成四维的集合．这是从其结构的观点看来的四维的集合，舍弃了各个现象的性质而且首先考虑相对论．它并不是以快速运动的理论、宇宙论、新的空间或时间的理论作为其基础和本质的，它只不过是把空间和时间看作物质存在的统一形式的理论罢了．

当然，在牛顿的力学里也可以把空间和时间统一在一个四维流形之中．我们已经提到过，多维空间的思想首先是拉格朗日考虑质点的运动时产生的，他在空间坐标 x、y、z 之外添上了时间 t．于是点的运动就由具有坐标 x、y、z、t 的四维空间中的曲线表出；在点运动时，位置 (x, y, z) 和时间 t 这四个坐标都改变．然而空间和时间的这种统一具有纯形式的特点，在这里并未在空间和时间之间建立任何内在的必要的联系．当然，在运动规律里，每一个物

体都有其空间位置对时间的依赖性．但是这牵涉的仅是每个已知的运动，不论是在力学里还是一般地在物理学里，在相对论以前都没有建立过空间和时间之间的任何一般的内在联系，总是单独地把事物和现象的空间关系、空间顺序与它们在时间里的关系和顺序区别开来．事物的时间的先后、时间间隔的延续性，都被认为是与什么东西也无关系地绝对的和确定的．简短地说，绝对的时间的概念在物理学中到处占统治地位．

爱因斯坦的伟大的发现，不仅形成了相对论的基石，而且是在空间和时间问题的一般物理和哲学认识里的转折点．这个发现指出，绝对的时间在现实世界中是不存在的．当爱因斯坦在 1905 年建立了他的理论[1]以后不久，闵科夫斯基指出，相对论的本质与其说是拒绝了绝对的时间，还不如说是确立了空间和时间的相互联系，因而就有了物质存在的统一的绝对形式：时空．而空间——空间坐标——从时间、从时间坐标 t 分出，在某种程度下是相对的，依赖于那种物质体系——“计算体系”的，现象的空间的和时间的顺序就是对于这体系而决定的．对于一个体系说是相同的事物，在对另一个体系来说时可以不是相同的．

在决定现象的顺序时，当然不会也不能有与计算体系的完全的依赖性．由直接的相互作用联系的事物的顺序自然对于所有的体系说都是同一个样子，因而作用总在它的结果之前．但是对于并不与相互作用联系的事物，在时间方面的顺序就显得是相对的了．因为空间的顺序(在其纯粹的形式里)是对于同时的事物说的，而同时性却是相对的，所以把纯粹的空间关系从时空关系的一般集合中分出来显得是相对的，依赖于计算体系的．在绝对意义下的空间显得似乎是四维时空流形被同时的(对于已知体系而说)事物所作的截面．

在我们的问题里不可能引入相对论原理的叙述，我们只能尽量用最简短的话以这样的形式来刻画它的原理，使得它们能够在

1) 爱因斯坦在 1905 年建立的理论叫做狭义相对论，这是对于爱因斯坦在 1915 年建立的“广义”相对论说的．

与抽象几何的思想的联系中最自然地被理解. 顺便说一下，这个理解与爱因斯坦自己的出发点也有实质的区别.

宇宙可以看作是各种各样的事物的集合. 所谓事物，在这里不仅是指在空间中伸展而且在时间中延续的现象，而且是指任何瞬时而又点状的事物，类似于点状灯火的瞬时爆发. 用几何的语言来说，事物是宇宙的四维流形的点.

时空是物质存在的形式，是这个宇宙流形的形式. 时空的结构，它的"几何"，不是别的，而是宇宙的某个一般的结构，即根据我们的讨论是事物集合的"几何". 这个结构由事物的某些普遍的物质的相互联系和关系决定.

首先，我们在上面已经对于空间关系解说过，这应该就是物质的关系和相互联系. 对于现象在时间里的关系这也是对的. 以纯粹形式出现的空间的和时间的关系本身只是抽象罢了.

其次，决定时空结构的事物的关系，应该根据时空的普遍性而有普遍的特性.

事物的这种普遍的物质关系表示它们的因果联系. 每一个事物这样或那样地、直接或间接地作用于其他一些事物，而它自己又经受其他事物的作用. 一些事物作用于另一些事物的这个关系就决定了时空的结构.

因此，相对论许可给出下列定义. 时空是事物的集合，这时我们舍弃了事物的所有具体的性质和关系，只保留一些事物作用于另一些事物的普遍关系. 这里在一般的意义下，这个关系本身也应该在从它的各种各样的具体形式的抽象中来理解.

在狭义相对论里，时空算是最大限度均匀的. 这表明事物的流形许可作不破坏事物间的作用关系的变换，而且这种变换的群在一定的意义下是最大可能的. 例如设有两对事物 A, B 和 A', B'，而且 A 不作用于 B，B 也不作用于 A；对于 A' 和 B' 也有同样的情况. 于是就存在事物之间的这种对应，它使 A 对应于 A'，B 对应于 B'，而且这时对于任何一对事物，作用(或者不作用)关系不被破坏，即如果 X 作用于 Y，则对应的事物 X' 也作用于

Y'，而且如果 X 不作用于 Y，则 X' 也不作用于 Y'.

由此可见，从狭义相对论的观点看来，时空是一种四维空间，它的几何由某个变换群决定．这种变换不是别的，正是著名的罗仑兹变换．几何和物理的规律在这种变换下不改变．这种看法对应于在§6里说过的、由克莱因在他的爱耳浪根纲领里提出的那种对几何学的看法．

6．广义相对论跑得还要远些，它放弃了关于时空的均匀性的观念．它认为时空只是在充分小的区域里以一定的近似性而均匀的，但是整个却是不均匀的，根据爱因斯坦的理论，时空的不均匀性由物质的分布和运动决定．时空的结构本身又决定物体的运动规律，而且这是在万有引力的现象里发现的．广义相对论就其本质说是在时空结构与物质运动的联系中说明引力的引力理论．

只在微小的区域内以一定的近似性而均匀的时空的观念，类似于只"在无穷小范围内"是欧几里得式的黎曼空间的观念．从数学上看，当然在实质上有改变的意义下，广义相对论里的时空可以解释成一种黎曼空间．

那就是说，在四维黎曼空间里，在每个点邻近可以引进坐标，使得线性元素的平方由公式 $ds^2=dx_1^2+dx_2^2+dx_3^2+dx_4^2$ 表出．

在时空里，可以在每个事物的邻域里引进坐标，使得线性元素由公式 $ds^2=dx^2+dy^2+dz^2-c^2dt^2$ 表出，这里 c 是光速，它在测量单位的对应选取下可以认为就等于单位数．公式中的 x, y, z 是空间的坐标，t 是时间．dt^2 前的负号形式地表示了时间坐标与空间坐标（时间与空间）的实质的、根本的区别．

在引力理论里起最重要作用的是关于时空的曲率的概念．由爱因斯坦给出的这个理论的基本的方程，正好联系了刻画时空的曲率的量与刻划物质的分布和运动的量．这些方程同时是引力场的方程，而且正像爱因斯坦在佛克的合作下所证明的，从其中引出物体在引力场中的运动规律．

按照广义相对论，时空的结构显得是复杂的，而且即使在狭义相对论里可以允许的那种程度下，空间也不能从时间分出．然而

以一定的近似性和在一定的假定下这是可以做到的．在与宇宙尺度相比是微小的区域里，空间以充分的正确性显得是欧几里得式的，但是在较大的区域里却显示出与欧几里得几何的差异．这个差异依赖于物质质量的分布和运动，而且在巨大质量的星体邻近还达到值得注意的(虽然仍然是很小的)量，而整个地都在宇宙尺度之内．在一系列关于宇宙结构的假设中，整个说来都近似地假定质量是均匀地平均分布的．在一个这种假定的理论中，由苏联物理学家兼数学家弗利德曼假定的空间的几何整个地重合于罗巴切夫斯基几何．

在相对论里，抽象几何不仅是作为数学工具而得到其应用，抽象空间的观点本身也给出了这个理论原理的最深入的表述方法．在现实中开辟了由抽象几何来规划的可能性，而且理论上的思维获得了辉煌的胜利．从物体的空间关系和形式的经验研究生长起来的抽象几何，现在与作为现成的数学方法的现实空间的研究相对立了．这是科学的一般途径：它从经验中的直接数据提高到理论上的推广和抽象，它们重又回到经验，作为更深入地认识现象本质的工具，给出已知现象的解释和新现象的预言，指出人类实践活动的方向，而且在其中求得自己的证明和进一步发展的泉源．

文　　献[1]

А. Д. Александров, Геометрия. БСЭ, т. стр. 533—550.

Б. Н. Делоне, Краткое изложение доказательства непротиворечивости планиметрии Лобачевского. Изд. 2 е, Гостехиздат, 1956.

П. А. Широков и В. Ф. Каган, Строение неэвклидовой геометрии, Гостехиздат, 1950.

包含罗巴切夫斯基几何的内容，它的叙述易懂并且比较完全．

В. Ф. Каган, Геометрические идеи Римана и их современное развитие. ГТТИ, 1933.

Д. И. Блохинцев и С. И. Драбкина, Теория относительности А. Эйнштейна, Гостехиздат, 1940.

В. А. Фок, Современная теория пространства и времени. “Природа”, №12, 1953.

1) 还请参看第七章(第二卷)和第十八章的文献．

以尽可能易懂的形式讲述关于现实空间的几何学的现代观点原理的文章.

В. А. Фок, Теория пространства, времени и тяготения, Гостехиздат, 1955.

Н. В. 叶非莫夫: 高等几何学(分上、下两册),高等教育出版社, 1956.

П. К. 拉舍夫斯基: 黎曼几何与张量分析,高等教育出版社, 上册 1955, 下册 1958.

Э. Картан, Геометрия римановых пространств, ОНТИ, 1936.

为受过训练的读者广泛地讲述对象的专门著作.

裘光明 译

吴祖基 校

第十八章 拓 扑 学

§1. 拓扑学的对象

“附贴性是物体的一个特殊的属性. 如果我们把这个性质掌握, 而把物体其他的一切属性, 不问是本质的或偶然出现的, 均不予考虑, 那么就说所考虑的是几何物体.”

这是罗巴切夫斯基在他的著作“新几何原本”[1] 第一章开头的一句话.

当用图 (图 1) 来说明上面这句话的意思时, 罗巴切夫斯基继续写道: “两个物体 A, B 互相粘合时产生几何物体 C……; 反之, 每个物体 C 被任意的截面 S 划分为两部分 A, B”.

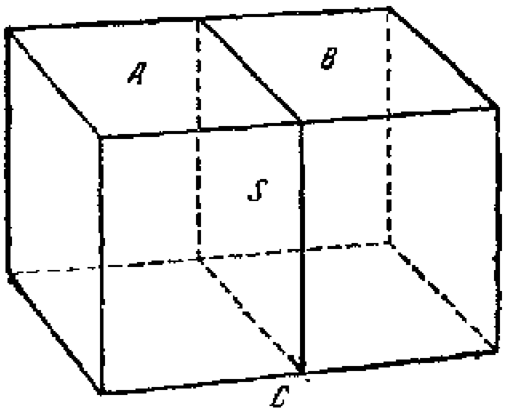


图 1

这些关于附贴、近傍、无限邻近的概念, 以及在某种意义下与之对偶的物体的截断概念被罗巴切夫斯基拿来当作整个几何学的基础概念. 如果用现代眼光来看的话, 实质上这也是拓扑学中基本的且占最主要地位的概念. 因此, 这位伟大几何学家的近代评论者们说: “罗巴切夫斯基在数学史上第一次试图从物体的拓扑性质出发来构筑几何学……[2]、罗巴切夫斯基用物体附贴与截断的术语来决定曲面、曲线和点的概念”, 他们这样的评论是完全正确的, 足以反映出附贴与截断概念

1) Н. И. 罗巴切夫斯基全集, 第二卷, 国立科学技术理论书籍出版社, 1949, 168 页.

2) “新几何原本”注解, 同书, 465 页.

的一些含有多样性内容的具体几何图像，这些可以在下列的图2里看到．这些图例是罗巴切夫斯基为着同样的目的而举出的，我们取自上面引到过的原书．

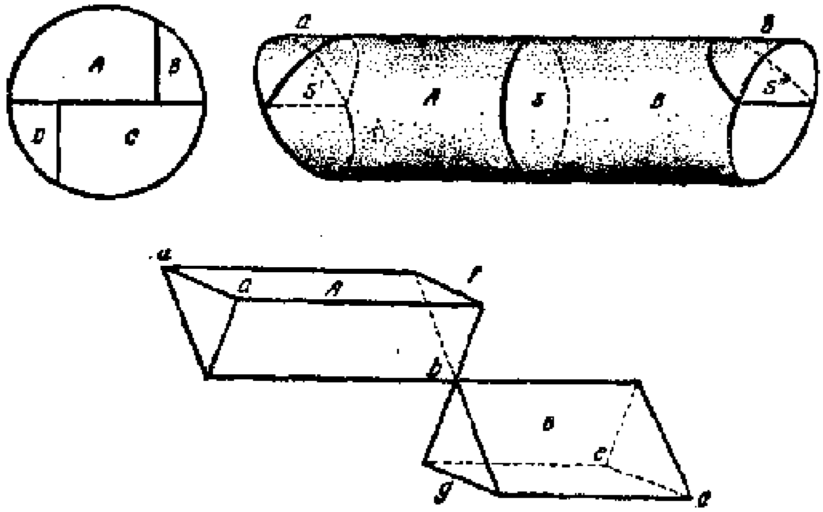


图 2

如果几何图形的变换不破坏图形各部分的附贴关系，则这个变换叫作连续变换；如果不仅仅附贴关系没有被破坏，并且也没有增添出新的附贴关系，则变换叫作是拓扑变换．因此，任何图形受到拓扑变换时，图形中互相附贴的部分仍然是互相附贴的，而不互相附贴的部分，仍然是不互相附贴的．简单地说，在拓扑变换之下既不会产生断裂，也不会产生粘合．特别，两个不同的点不能合为一点（在这个情形产生了新的附贴，图3）．因此，如果把图形看作是由它的点所产生的，则拓扑变换不仅仅是连续的，并且是一点对一点的对应：图形中任意两个不同的点变为两个不同的点．于是，

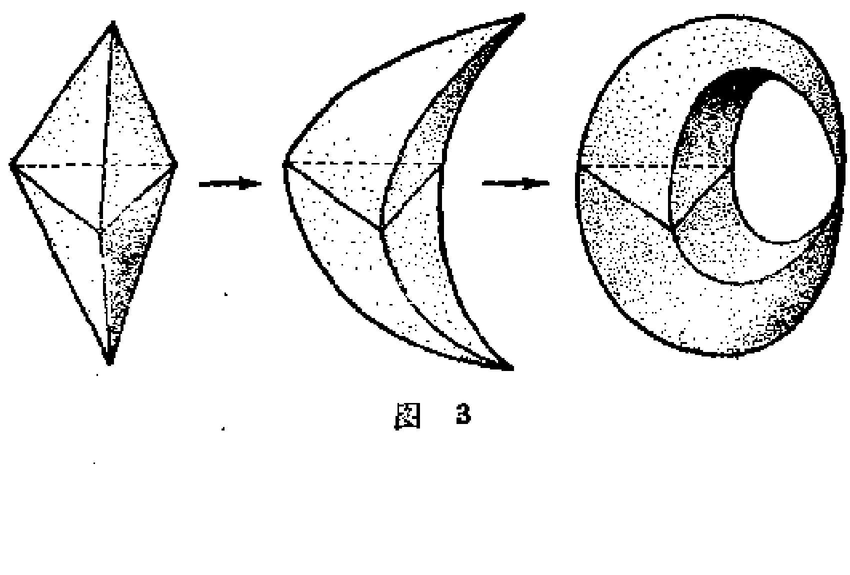


图 3

拓扑变换就是双边连续与双边一对一的变换.

从直观上来看，任意一个几何图形（曲线、曲面等等）的拓扑变换可以用下列方式来表示.

假设我们的图形是用某种柔韧的、可伸缩的物质，例如橡皮造成的，于是可以使图形作各种可能的连续形变. 这时，图形在它的某些部分扩伸了，而在另一些部分压缩了. 一般地说，大小与形状都将改变. 例如，我们可以把一个圆形的橡皮圈拉长为一个极其狭长的椭圆，也可以把它变为正多边形或非正多边形的形状，更可以把它变为形状非常奇怪的闭曲线. 图 4 里画出了一些，但我们不能够用拓扑变换把圆圈变成 8 字形曲线（要这样作，必须把两个点粘合，图 5），或者线段（要达到这个目的，可以把圆周上的一个半圆与另一个半圆粘合，或者把圆圈在一点截断）. 圆周是简单闭曲线，它只含有一环，与 8 字形曲线含有两环以及三瓣曲线（图 5）含有三环不同. 圆周是简单闭曲线这一性质在拓扑变换之下是保持不变的，是所谓拓扑性质.

如果我们取一个球面来，譬如小橡皮球，则可以用拓扑变换来使它的形状大有改变（图 6）. 但我们不能够用拓扑变换把球面变

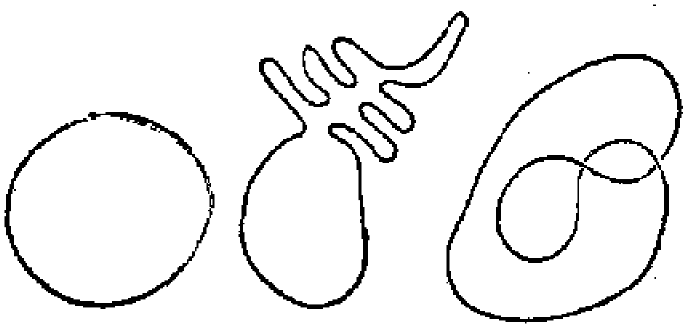


图 4

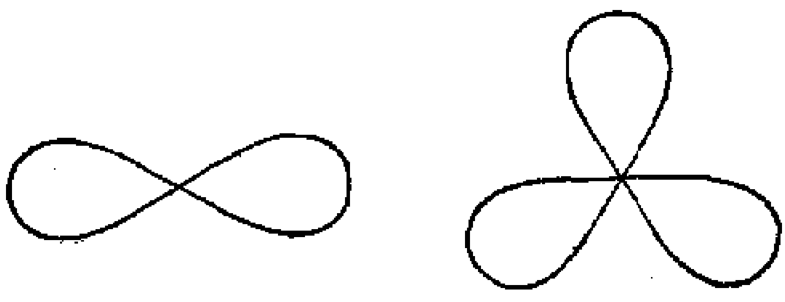


图 5

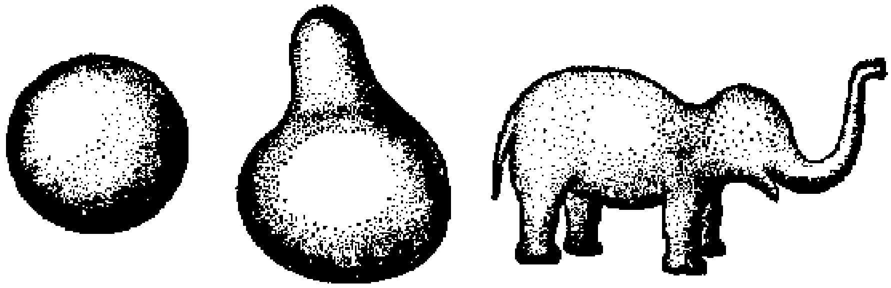


图 6

为正方形或者环面(面包圈的表面或救生圈那样的曲面,图 7). 事实上,球面具有下列两个在任何拓扑变换之下保持不变的性质. 第一个性质: 球面是闭曲面, 它没有边界(而正方形有边界); 第二个性质: 球面上每一条闭曲线,用罗巴切夫斯基的话来说,是它的截线. 如果沿着一条闭曲线切开我们的小橡皮球, 则球面将被分割成两块互不联接的部分. 环面则不具有这个性质: 如果把环面沿着它的一条子午线切开(图 7), 它不致于分成许多块,只不过是变成了一个弯曲的圆筒形(图 8), 而这个圆筒形又不难经过拓扑变换(拉直)变为圆柱面. 因此,与球面不同,在环面上并不是每条闭曲线都是截线. 因此, 球面不能拓扑地变为环面, 我们说, 球面与环面是拓扑上不同的曲面,或者说它们属于不同的拓扑类型,或者说这两个曲面互不同胚. 反之, 球面与椭球面以及一般的有界凸闭曲面属于同一个拓扑类型,也就是说,互相同胚. 这就是说,用拓扑变换可以把它们之中的一个变为另一个.

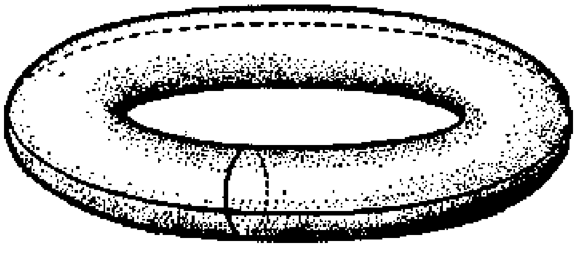


图 7

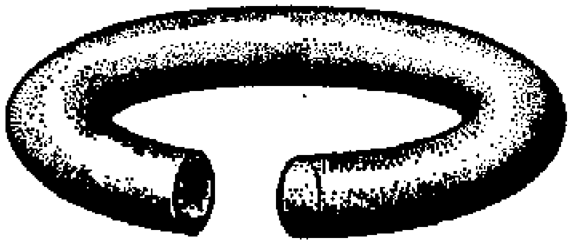


图 8

§2. 曲 面

上面已经说过, 几何图形在任何拓扑变换之下不变的性质叫

作拓扑性质. 拓扑学研究图形的拓扑性质,不仅如此,它还研究几何图形的拓扑变换以及任意的连续变换.

我们方才已举了一些拓扑性质的例子. 下面这些都是拓扑性质:曲线或曲面的闭性质,闭曲线是简单闭曲线的性质(也就是说,只含有一环),闭曲面被它上面任意闭曲线分割的性质(球面具有这个性质,而环面不具有这个性质),等等.

在曲面上引闭曲线,使得把它们放在一起不组成曲面的划分,也就是说,不把曲面分割为不连接的部分. 在一个曲面上所能够引的这样的闭曲线的最大数目,叫作这个曲面的连通阶. 这个数目将使我们得到对于曲面拓扑结构的最重要的知识. 我们已经看到,对于球面来说,它等于零(球面上的任何闭曲线组成划分). 在环面上可以找到两条闭曲线,它们合起来还不致于分割环面. 其中之一可取任意一条子午线,另一则取任意的平行环(图 7). 但是在环面上不存在三条共同不分割这个曲面的闭曲线. 环面的连通阶等于 2. 像图 9 里的面包,它的表面就是一个连通阶等于 4 的闭曲面,等等. 一般地说,在一个球面上开 $2p$ 个小圆洞(图 10 上是 $p=3$ 的情形),而把这些圆洞配为 p 对,在每一对上粘一个圆柱面(沿着边粘),则得到

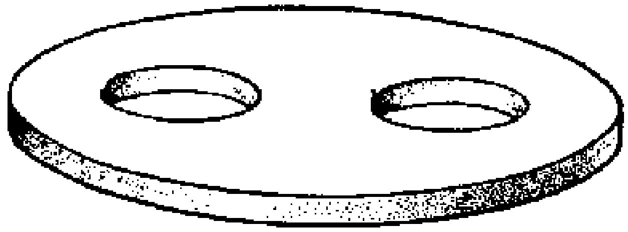


图 9

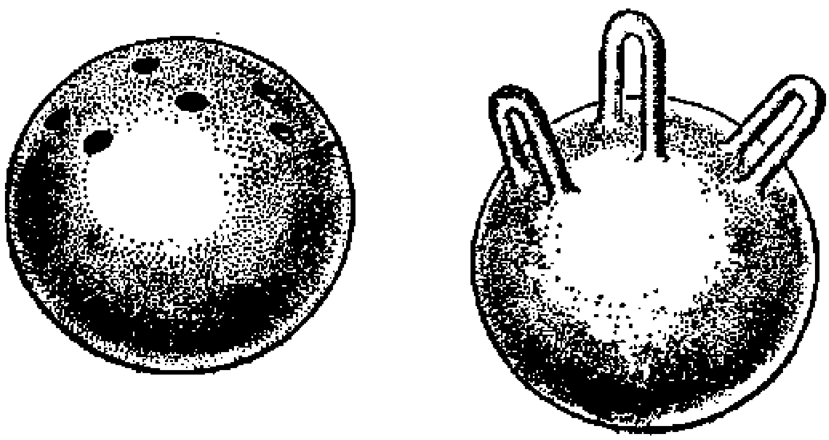


图 10

一个手柄．这样我们就得到了一个带有 p 个手柄的球面，或者叫作亏格为 p 的规范曲面．这个曲面的连通阶等于 $2p$．

用罗巴切夫斯基的说法，所有这些曲面都是空间的“截面”：它们每一个都把空间分为两个区域，内部与外部，而它们自己是这两个区域的公共边界．这个性质又与另一个性质有关，那就是这些曲面当中的每一个都具有两侧：外侧与内侧（其中一侧可以用一种颜色涂满，而另一侧则可用另一种颜色来涂）．

但是与这些曲面并存的还有所谓单侧曲面，对它们来说无从区别两侧．最简单的例子就是熟知的“莫毕鄂斯带子”．把一个长方形纸条 $ABCD$ 的 AB 边与 CD 边粘合，但粘时使顶点 A 合于顶点 C，顶点 B 合于顶点 D，我们就得到如同图 11 所画的曲面．它叫作莫毕鄂斯带子．不难看出，在莫毕鄂斯带子上不能区分两侧，不能用不同的颜色来涂满两个侧面：沿着带子的中线运行，从点 E 开始运行一周以后再回到点 E 时，虽然不曾经过带子的边界，但已到了与原来出发时不同的一侧．顺便指出，莫毕鄂斯带子的边界是由单独一条闭曲线所组成．

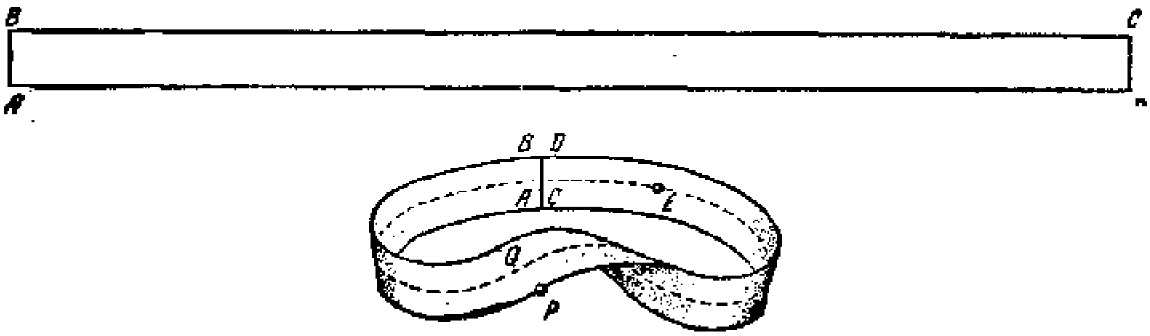


图 11

于是产生了这样的问题：是不是存在单侧闭曲面，也就是说，不具有边界的单侧曲面？这种曲面是存在的，但如果把它们置放在三维空间里，则必然会发生自己与自己相交的情况．典型的单侧闭曲面的例子就如同图 12 里所画的，叫作“单侧环面”或克莱茵曲面．如果要避免自己交叉，设想把两个莫毕鄂斯带子沿着边界粘合起来便得到克莱茵曲面（如同上面曾经提到过的，莫毕鄂斯带子的边界只含有一条周线）．

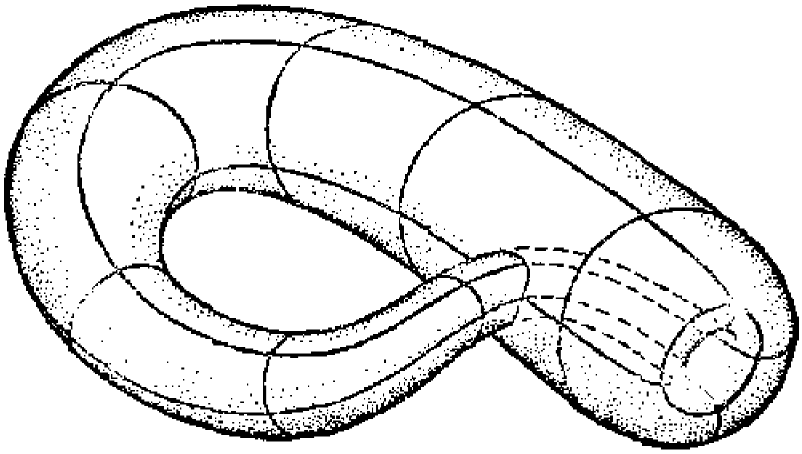


图 12

现在我们可以来陈述曲面拓扑学的基本定理应用于双侧曲面的情形：任何一个双侧闭曲面必同胚于某个亏格为 p 的规范曲面，也就是"具有 p 个手柄的球面"．两个双侧闭曲面同胚，当且仅当它们具有相同的亏格 p（相同的连通阶 $2p$），也就是说，它们同胚于具有同样多数目手柄的球面．

对于单侧曲面来说，也有"规范形式"，类似于亏格 p 的双侧曲面所有的规范形式，但是比较难以表现出来．要得到这种规范形式，取一个球面来，在上面开 p 个圆洞，在每个洞上粘一个莫毕鄂斯带子，使得带子的边界与圆洞口的边界粘合．试图作这种粘合时将碰到困难，因为根本不可能在空间里具体地实现．在施行这样的粘合时，必然又将导致曲面的自己交叉．而这种交叉是在空间里实现单侧闭曲面的模型时所不可避免的．

不能认为单侧闭曲面只是数学里的趣谈，而与严肃的科学问题无关．要证实这种想法的错误，只需以射影几何的产生为例证．这是几何思想的一大成就．这门学科的原理现在已被列入大学或师范学院的课程中．射影几何学的实际应用起源于配景理论，是在文艺复兴时期（列昂纳都·达·芬奇）由于建筑、绘画与技术制图的需要而产生．射影几何学最初一些定理的发现是十六—十七世纪的事，是由于纯粹实用上的需要而产生．射影几何的发展成为几何学在纯理论方面的一个最突出的推广．特别，在它的基础之上，

人们初次彻底地理解了罗巴切夫斯基的非欧几何学[1].

要从初等几何里所讨论的普通平面得到射影平面，只需在平面上补充一些抽象的元素，所谓非正常点或“无穷远点”．仅当作了这种补充以后，从一个平面到另一个平面的射影(从射影灯到银幕的射影)才是从一个平面到另一个的一对一变换．把平面补充以非正常点，在解析几何里就相当于从通常的笛卡儿坐标过渡到齐次坐标．也可以按照下述的方式来描写：每一条直线补充了唯一的一个非正常点(“无穷远点”)；两条直线具有相同的无穷远点当且仅当它们是平行的．添加了唯一的无穷远点以后，直线变成了闭曲线，而所有各直线上无穷远点的全体按定义组成非正常的或无穷远直线．

因为平行的直线具有共同的无穷远点，所以，要表达出在平面上补充无穷远点的实际步骤，只需考虑通过平面上某个定点的直线，例如通过坐标原点 O 的直线(图 13)．这些直线的无穷远点已经把平面上所有的无穷远点包罗无遗(因为每一条直线与通过点 O 的平行直线具有同一个无穷远点)．因此，我们得到了射影平面的一个“模型”．这个模型是在一个以 O 为中心，具有“无穷大”半径的圆里，把圆周上每一对对径点(同一条直径的两个端点)粘合为直线 AA' 唯一的“无穷远”点而得到的．整个圆周就变成了无穷远直线，但这时必须记牢，圆周上的每一对对径点是看作互相等同的一点的．由此立刻可以看出．射影平面是闭曲面，它没有边界．

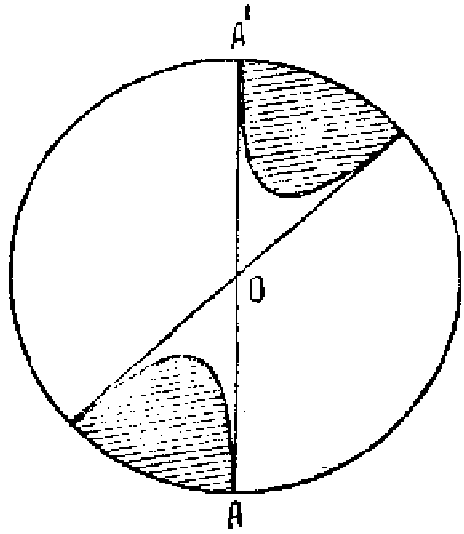


图 13

如果在射影平面上考虑一条形状如双曲线（见图 13）的二次曲线，则很显然这条双曲线在射影平面上是一条闭曲线(只不过是

1) 例如参照第十七章 § 6，或 П. С. 亚历山大洛夫的小册子“什么是非欧几何”(1951).

被无穷远直线分割成两部分）．既然我们取作基本圆的圆周上每一对对径点是粘合为一的，则不难看出，在图13上用阴影表示的双曲线内部同胚于普通的圆，而射影平面上没有阴影的其余部分则同胚于莫毕鄂斯带子．因此，从拓扑的观点来看，射影平面是把圆形（在我们的情形是双曲线内部）与莫毕鄂斯带子沿着边界粘合以后所得的结果．由这里就立刻推知，射影几何所研究的基本对象——射影平面，是一个单侧的闭曲面．

除了极其重要的几何意义以外，射影平面的有趣之点还在于它突出地表现了近代几何思想的一个特点，这种特点是基于罗巴切夫斯基的发现而形成的．由于几何图形这个概念的特征，几何的思想总是抽象的．现在，它又进一步抽象化，如同我们所说的，在普通平面上补充进新的抽象元素——非正常点来．当然，这些抽象元素也反映了现实实际中的事物（每个“非正常点”不过就是一束平行直线的抽象化），但它们在我们的考虑中只是一种抽象的几何元素，这种元素只能不完全地（不能予以物理实现地）用粘合某个圆周上对径点后的结果来表示．在整个近代拓扑学里，类似的抽象构造具有很大的价值，特别是当从曲面进入到三维或高维流形时，情况是如此．

§3. 流　　形

考虑下面的简单仪器，有时也叫作平面复摆（图14）．它是由两条轴 OA 与 AB 在点 A 用铰链联牢而构成的．点 O 是固定不动的，轴 OA 可以在一个固定的平面里围绕着点 O 而转动，而轴 AB 也是在那一个平面里可以绕着点 A 而自由转动．我们这个系统的每一个可能的位置可由两个角度，φ 与 ψ，完全决定，这里 φ 与 ψ 是轴 OA 与 AB 同平面里某个固定方向所成的角，例如，与横标轴的正方向所成的角．这两个在 O 与 2π 之间变动的角度可以看作环面上一点的“地理坐标”，这坐标是关于环面的“赤道”以及某一条“子午线”而取的（图15）．

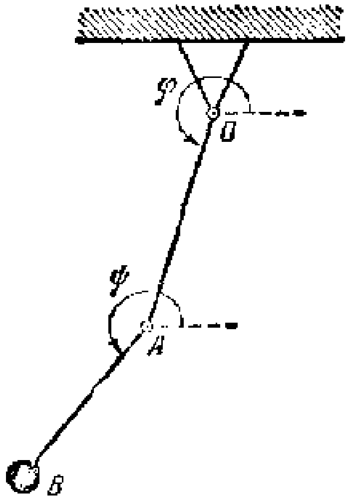


图 14

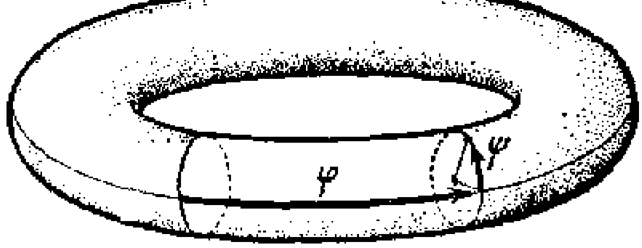


图 15

因此，可以说我们这个力学系统所有可能的状态所形成的集合是一个二维流形——环面．把每个角 φ 与 ψ 换为与之相当的、在某两个圆周上的点，在这些圆周上已确定了计算弧长的起点与方向(图 16)，则我们也可以说上述力学系统的每一个状态决定于在这两个圆周的每一个上取定一点(在一个圆周上取纬度 φ，在另一个上取径度 ψ)．换句话说，完全如同在解析几何里把平面上的点对应了一对数——坐标那样，在现在的情形我们把环面上的任意一点(因而平面复摆的任意位置)对应了它的一对地理坐标，也就是说，对应了一对点，其中每个各在一个圆周上．这件事实就表示说，平面复摆所有的状态所形成的流形，也就是说环面，是两个圆圈的拓扑乘积．

现在把我们的仪器作下述的改变：它仍然是由两条轴 OA 与 AB 组成，轴 OA 可以在某个固定的平面里围绕着点 O 而转动，AB

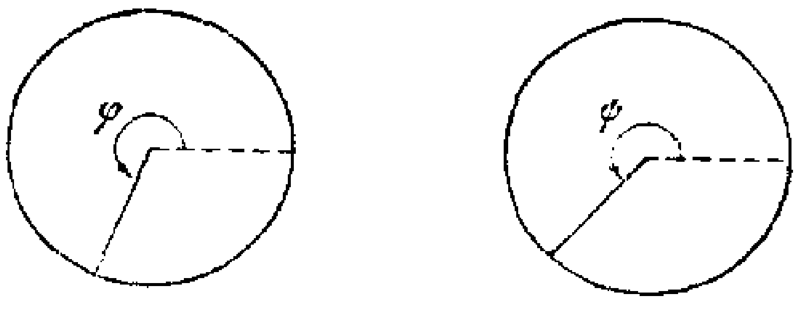


图 16

轴在点 A 与轴 OA 用铰链相接．但这时，无论点 A 在什么位置，轴 AB 可以围绕着点 A 在空间里自由转动，它可以取任何由点 A 出发的射线所取的方向．这时，我们的系统由三个参数决定，其中一个仍然是轴 OA 与横标轴正方向的交角，而其余的两个参数决定轴 AB 在空间里的方向．后两个参数可以这样来给出：以坐标原点 O 为中心引一个单位球面，取球面平行于轴 AB 的半径 OB'，则点 B' 或它在球面上的那两个地理坐标就可作为是决定轴 AB 方向的参数．于是，我们这个新的铰链系统所有一切可能的位置所形成的集合是一个三维流形．读者不难想到，这个流形可以解释作圆周与球面的拓扑乘积．这个流行是闭的，也就是说没有边界，因此，不能在三维空间里把它用图形具体地表示出来．若要把这个流形表达得比较直观一些，考虑空间里夹在两个同心球中间的部分．从它们的中心发出的每一条射线同它们各交于一点．如果把这样的每一对点等同起来（粘合为一点），我们就得到一个三维流形，正是圆周与球面的拓扑乘积．

我们可以把上述的仪器进一步复杂化，不仅仅使由铰链接合于轴 OA 上的轴 AB 可以自由地绕着点 A 在空间里转动，并且使轴 OA 也可以自由地绕着点 O 在空间里任意转动．这个系统所有可能的位置所形成的集合将是一个四维闭流形——两个球面的拓扑乘积．

我们看到，在考虑最简单的力学（运动学的）问题时，将引出拓扑流形，并且是三维或者更高维的．实际上，对于力学问题作较为详细的研讨时，具有较大意义的流形（一般说也是高维的），是所谓动力系统的相空间，其中不仅包括了所给的力学系统可能取的几何位置，并且还包含了系统中各个点的运动速度．我们只举最简单的例子．设有一点在圆周上以任意速度运动．这个系统的每一个状态可以由两个数据决定：每一时刻点在圆周上的位置与速度．这个系统所有的状态所形成的流形（相空间），显然是一个无穷圆柱面（圆周与直线的拓扑乘积）．

相空间的维数随着系统的自由度而增加．有些系统的力学特

征可以由它的相空间的拓扑性质表现出来．例如，系统的周期运动相当于相空间内的闭曲线．

对于由力学、物理学、天文学（天体力学，宇宙学）的问题所提出的动力系统的相空间而作的研讨，使得数学家注意到高维流形的拓扑学．正是因为这些问题才使得伟大的法国数学家庞加莱在上个世纪九十年代运用所谓组合方法来系统地建立了流形的拓扑学，而这个组合方法至今仍然是拓扑学中的一个基本的方法．

§4．组合方法

在历史上，关于拓扑学的第一个定理是欧拉定理或欧拉公式（似乎笛卡儿就已经知道这个结果）．定理的内容如下：取任意一个凸多面体的表面来，令 α_0 表示它的顶点的数目，α_1 表示棱的数目，α_2 表示面的数目，则下面的所谓**欧拉公式**成立：

$$\alpha_0-\alpha_1+\alpha_2=2. \tag{1}$$

显然，当凸多面体经受到任意的拓扑变换时，公式仍然成立．因此这个几何定理是属于拓扑学的．当然，经过拓扑变换以后，棱不一定仍然是直的，面也不见得仍然是平的，不过对于弯曲面的数目、弯曲棱的数目以及顶点数目之间的关系(1)仍旧保持有效．最重要的情形是各个面都是三角形，这时叫作一个**三角剖分**（把我们的曲面剖分成直的或弯的三角形）．很容易把一般多面体的情形化为这个情形：只需把各个面剖分为三角形（引每个面的一些对角线）．这样，我们就可以限于三角剖分的情形．曲面拓扑学的组合方法就在于把对于曲面的研究用对于它的三角剖分的研究来代替，当然我们所感兴趣的只是不依赖于三角剖分的选取的那种性质．既然是已给曲面的一切三角剖分所共有的性质，因此也将表示曲面本身的某种性质．

欧拉公式提供给我们这样的性质之一，现在我们详细地考虑它．欧拉公式的左方，也就是式子 $\alpha_0-\alpha_1+\alpha_2$ 叫作这个三角剖分的**欧拉特征数**，其中 α_0 是这个三角剖分的顶点总数，α_1 是棱的总

数，而 α_2 是三角形的总数．欧拉定理说，对于同胚于球面的曲面，它们的任何三角剖分都具有等于 2 的欧拉特征数．对于任何曲面(不仅是同胚于球面的曲面)来说，它们的任何三角剖分都具有相同的欧拉特征数．

要想知道各种曲面欧拉特征数的数值也不困难．首先，对于圆柱面来说，这个数值为零．事实上，从球面的任意一个三角剖分里挖去两个不相邻接的三角形，但把它们的边仍旧留下，则显然我们得到同胚于圆柱侧面的曲面以及它的一种三角剖分．这时，顶点与棱的数目没有变动，而三角形的总数少了两个，因此，这个三角剖分的欧拉特征数等于零．现在，从球面的某个三角剖分里挖去 $2p$ 个两两互不相邻(没有公共边，也没有公共顶点)的三角形[1]，这时，欧拉特征数将比原来减少 $2p$．不难看出，如果把球面上这 $2p$ 个洞口分为 p 对，在每一对洞口粘上一个弯曲的棱柱形手柄，则欧拉特征数仍然不变．这是由于我们已看到棱柱形手柄的欧拉特征数等于零，而在柱形的边上，棱的总数又与顶点的总数相同．于是，亏格为 p 的双侧闭曲面具有等于 $2-2p$ 的欧拉特征数(这个事实首先是法国海军上将德·容克艾证明的)．

我们再引进三角剖分的一个重要性质，它也满足所谓拓扑不变的性质(也就是说，如果曲面的某一个三角剖分具有这种性质，则该曲面的任何三角剖分也都具有这种性质)．这个性质叫作可定向性．在没有陈述它的意义之前，注意每一个三角形是可以定向的，也就是说，在它的边界上确定一个环行的方向．三角形的两种可能的定向可以由它顶点的排

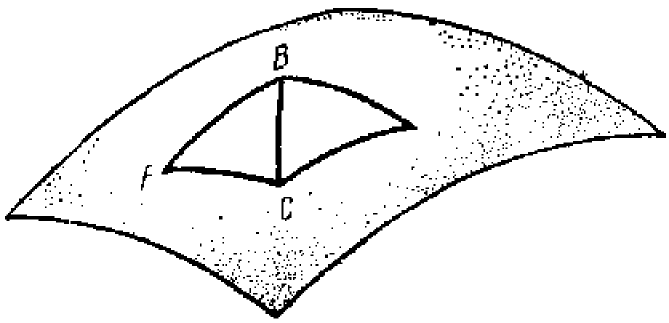


图 17

1) 要使得能这样作，只需使所考虑的三角剖分是适当地"细小"．任意一个三角剖分如果还不满足这个要求，只需把它再重分若干次便可以使它合用．

列顺序来决定[1]. 设在某个曲面上已给两个相邻接的三角形，它们有一条公共边，但没有其他的公共点(图 17). 如果这两个三角形所取定的定向在公共边上所诱导的定向相反，则我们说这两个三角形上所取的定向是**相符的**(在平面上或在任意双侧曲面上，这表示说，如果从曲面的某一侧来观察曲面，这两个三角形的周界按照相同的方向而环行，也就是说或者都是顺时针方向，或者都是逆时针方向). 如果能够把闭曲面某个三角剖分里所有的三角形都给以定向，使得每两个具有公共边的三角形都具有相符的定向，则这个三角剖分叫作可定向的. 下列的事实成立：双侧曲面的任何三角剖分是可定向的，单侧曲面的任何三角剖分是不可定向的. 因此，双侧曲面也叫作可定向曲面，单侧曲面也叫作不可定向曲面. 取莫毕鄂斯带子的任意一个三角剖分来，读者不难看出它的不可定向性. 若要得到射影平面的最简单的三角剖分，只需在它上面任意引三条不交于一点的直线(图 18). 它们把射影平面剖分为四个三角形，其中一个位于平面的有限部分，而其他三个之中的每一个都被无穷远直线割断. 在图 18 里，把通向无穷远的三个三角形之一用阴影描出. 在这个图上我们也就看出，任何把这四个三角形相符地予以定向的企图都是失败的. 特别，如果采用图 18 上对四个三角形所取定的定向，则它们的边界的代数和将是二倍的直线 AB，而不是零；但在相符的定向时，这个代数和应当为零.

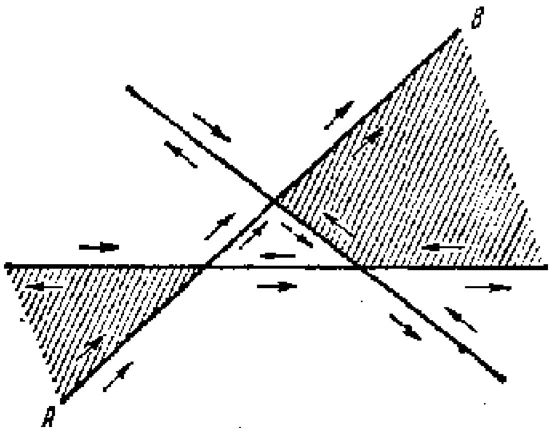


图 18

欧拉特征数以及可定向或不可定向性给出了关于闭曲面的所

1) 不难看出，顶点的两种排列顺序决定同一个定向(同一个环行的方向)当且仅当它们之中的每一个可以经过一个"偶"置换而变为另一个. 例如，(ABC)，(BCA)，(CAB)决定三角形的一种定向，而(BAC)，(ACB)，(CBA)决定另一种. (关于置换的奇偶性，参照例如第二十章 § 3.)

谓拓扑不变量的一个完全系. 意思是说，两个曲面同胚当且仅当它们的三角剖分具有相同的欧拉特征数，并且同为可定向或不可定向的.

组合方法不仅仅可用来研究曲面(二维流形)，并且可以用来研究任何维数的流形. 例如，在三维流形的情形，相当于剖分为三角形的将是剖分为四面体，叫作三维的三角剖分或流形的单纯剖分. 三维三角剖分的欧拉特征数是了解为和数 $\alpha_0-\alpha_1+\alpha_2-\alpha_3$，这里 $\alpha_i(i=0, 1, 2, 3)$ 是这个三角剖分里 i 维元素的总数(也就是说，α_0 是顶点数，α_1 是棱数，α_2 是二维面数，α_3 是四面体数). 当维数 $n>3$ 时，则把流形剖分为 n 维单纯形，也就是说，最简单的 n 维凸多面体，类似于三角形($n=2$)与四面体($n=3$). n 维流形所剖分成的 n 维单纯形以及这些单纯形的各维面(也是单纯形)共同组成这个流形的三角剖分. 这时仍然可以讲欧拉特征数，这时就了解为和数 $\sum_{i=0}^{n}(-1)^i\alpha_i$，这里 α_i 是在三角剖分中所出现的 i 维元素的总数 ($i=0, 1, 2, \cdots, n$). 与以前一样，对于已给 n 维流形(以及一切与它同胚的流形)的任何三角剖分来说，所得到的欧拉特征数是同一的，也就是说，欧拉特征数是拓扑不变量. 至于拓扑不变量的完全系(就如同欧拉特征数及可定向性对于曲面那样)将是怎样的，就目前我们的知识水平来说，即使对于三维流形也还不能设想.

组合方法对于近代拓扑学的价值是巨大的. 这个方法开辟了应用某些代数方法来解决拓扑问题的可能性. 精明的读者也许当前面在射影平面的三角剖分里讲到定向三角形边界的代数和时就已看到向代数过渡的可能性. 事实上，如果三角形已定向，也就是说，确定了一种环行的方向，则它的边缘将自然地取作它所有的、赋予一定定向后的边，而每一边上的定向也正是按三角形的定向环行周界时所诱导的.

考虑在某个曲面的三角剖分里出现的全体三角形 T_i^2, $i=1, 2, \cdots, \alpha_2$. 它们之中的每一个可以有两种定向. 三角形 T_i^2 取定

可能的定向之一种以后记作 t_i^2，而同一个三角形取另一种(相反的)定向时则记作 $-t_i^2$. 同样，对于上述三角剖分中出现的全体一维元素(棱) $T_k^1(k=1, 2, \cdots, \alpha_1)$ 也可以定向，就是，取定它的两种可能的方向之一. 线段 T_k^1 取定某个方向时记作 t_k^1，则取定另一个方向时记作 $-t_k^1$. 于是，如果三角形 T_i^2 的边是线段 T_1^1, T_2^1, T_3^1，则定向三角形 t_i^2 的边缘将是赋予一定定向后的这三个边，也就是说，t_i^2 的边缘是由有向线段 $\varepsilon_1 t_1^1$, $\varepsilon_2 t_2^1$, $\varepsilon_3 t_3^1$ 组成，这里，如果棱 T_i^1 的方向与 t_i^1 的方向一致，则 $\varepsilon_i=+1$，否则 $\varepsilon_i=-1$. t_i^2 的边缘记作 Δt_i^2. 我们看到，$\Delta t_i^2=\sum \varepsilon_k t_k^1$，但这个和式可以了解为展开在三角剖分所有的棱上而取的，把那些不在三角形 T_i^2 的边上出现的棱 T_k^1 所对应的系数 ε_k 取作零.

很自然地可以进而考虑展开在三角剖分全体棱上的一般和式 $x^1=\sum a_k t_k^1$[1]. 这种和式的几何意义非常简单：实际上和式中的每一项是取自三角剖分的某个线段，但附加了一定的方向与一定的系数(一定的"重数"). 整个的代数和表示一条由线段所组成的途径，在这条途径里每一个线段所出现的次数就如同它前面的系数. 例如，若我们先按箭头所指的方向走完多边形 $ABCDEF$(图 19)，然后过线段 AA' 到多边形 $A'B'C'D'E'F'$ 上按箭头所指的方向走一圈，再后，通过线段 AA' 回到多边形 $ABCDEF$，按箭头所指的方向再走一圈，则我们将得到一个和式，其中线段 AB, BC, CD, DE, EF, FA 的系数是 2，线段 $A'B'$, $B'C'$, $C'D'$, $D'E'$, $E'F'$, $F'A'$ 的系数是 1，而线段 AA' 根本不出现(它的系数为零，因为在它上面按相对方向各跑了一次).

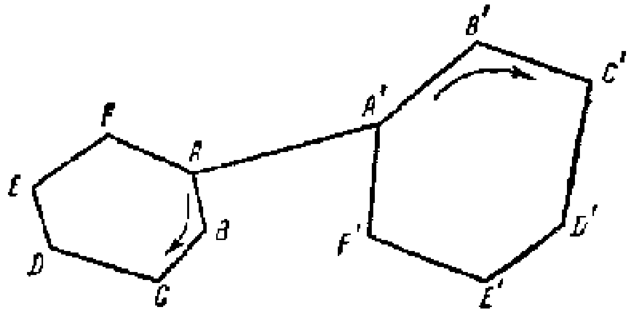


图 19

形式如 $x^1=\sum a_k t_k^1$ 的和式叫作三角剖分的一维链. 从代数的观点来看，它们是线性形式(一次的齐次多项式)，因而可以按照通

1) 系数 a_k 是整数。

常的代数运算法则相加、相减以及乘上任意的整数. 在一维链中特别重要的是所谓一维闭链. 在几何上,它们表示闭途径(在图19里所牵涉到的正是这种闭途径).

为要得到闭链的纯代数的定义,对于有向线段 $\overrightarrow{AB}$ 来说,我们令终点 B 在线段 $\overrightarrow{AB}$ 的边缘上带正号出现(以 $+1$ 为系数),起点 A 以负号出现(以 -1 为系数),则线段 $\overrightarrow{AB}$ 的**边缘**可以写为 $\Delta(\overrightarrow{AB})=B-A$.

有了以上的了解以后,我们立刻可以注意到(在通常意义之下)组成闭途径的线段 $\overrightarrow{AB}$, $\overrightarrow{BC}$, $\overrightarrow{CD}$, $\cdots$, $\overrightarrow{FA}$ 的边缘之和等于零. 这就自然地引出一般一维闭链的定义: 一维链 $x^1=\sum_k a_k t_k^1$ 是闭链,假如它每项的边缘之和为零, 也就是 $\sum_k a_k \Delta t_k^1=0$. 不难验证,两个闭链的和仍然是闭链. 闭链看作代数式而乘上任何整数以后仍然是闭链. 这就使得我们能够说闭链 z_1^1, z_2^1, $\cdots$, z_s^1 的线性组合就是形如 $z=\sum_{\nu=1} c_\nu z_\nu^1$ 的闭链,这里 c_ν 是整数.

类似于三角剖分上一维链的概念,可以讲这个三角剖分的二维链,也就是,形如 $x^2=\sum_i a_i t_i^2$ 的式子,其中 t_i^2 是三角剖分里的定向三角形. 因为每个定向三角形的边缘是一维闭链,所以链 $\sum_i a_i \Delta t_i^2$ 是闭链. 我们正是把这个闭链当作链 $x^2=\sum_i a_i t_i^2$ 的边缘 Δx^2.

利用**链的边缘**这个概念可以进一步陈述同调的概念: 三角剖分里的闭链 z^1 叫作在这个三角剖分里**同调于零**,假如它是这个三角剖分里某个二维链的边缘,在一切凸闭曲面的三角剖分里,或更一般些说,在同胚于球面的一切曲面的三角剖分里,任何一维闭链同调于零. 从几何上来看,这是很清楚的: 凸闭曲面上的每一个闭多边形必然是某一块曲面的边界. 但对于环面来说则不然: 环面上的子午线以及赤道都不是这个曲面上任何一块区域的边界. 如果取环面的任意一个三角剖分,则在这个三角剖分里可以找到类似于子午线与赤道的闭链,这种闭链不同调于零.

在前面所作的射影平面的三角剖分里，我们将看到一种新现象．如果把直线，例如直线 AB（图 18）看作是这个三角剖分里的闭链，则它并不同调于零．但是，同一条直条，如果取 2 为系数，则同调于零．因此，在链的定义里引入异于 ±1 的整数作系数，最初看起来似乎是形式的、不必要的，但现在就知道并不如此．这样作能够使我们抓着曲面以及一般流形的重要几何性质，在现在这个情形涉及到的是所谓挠系数的性质，就是在流形上存在那样的闭链，它本身不同调于零（不是某一块面积的边缘），但如果乘以某个适当的整数系数以后，就同调于零．

最后，与上面所说有关的一个极其重要的概念是闭链的同调独立性（或同调无关性）．闭链 $z_1, \cdots, z_s$ 叫作在三角剖分里是同调独立的，假如它们的任意线性组合 $\sum c_i z_i$ 仅当所有的 c_i 都为零时方才在这个三角剖分上同调于零．在环面上，作为同调独立闭链的例子，可以取任意一条子午线及赤道，把它们看作环面的某个三角剖分上的闭链．

整个组合拓扑学的基本概念——边缘、闭链、同调的概念——我们上面只是在一维的情形给了定义，但是这些定义可以逐字逐句地推广到任何维数．例如，二维链 $z^2=\sum a_i t_i^2$ 叫作是一个闭链，假如它的边缘 $\Delta z^2=\sum a_i \Delta t_i^2$ 等于零．三维链是形如 $x^3=\sum a_h t_h^3$ 的式子，其中 t_h^3 是定向的三维单纯形（四面体）．

如同在三角形的情形一样，三维单纯形（四面体）的定向可以由顶点排列的顺序来给定．顶点的两种顺序排列如果能够用一个偶置换互相变换，则决定同一个定向．三维定向单纯形 $t^3=(ABCD)$ 的边缘是二维链（闭链）$\Delta t^3=(BCD)-(ACD)+(ABD)-(ABC)$（图 20）．三维链的边缘就定义作它的每个单纯形的边缘冠以这个单纯形在该三维链里出现的系数以后所作的和式．读

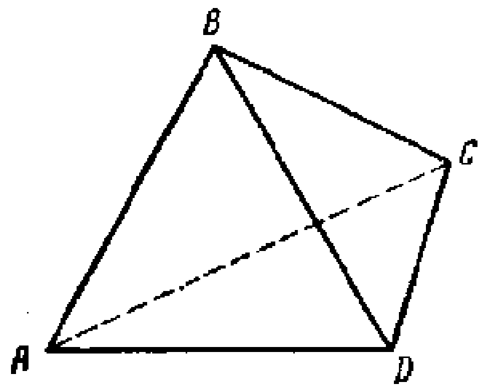


图 20

者不难验证，任何一个三维链的边缘是二维闭链(这只需对于三维单纯形来加以证明就够了)．我们说，二维闭链在流形里同调于零，假如它是流形里某个三维链的边缘．如此类推．注意，根据前面所说可定向与不可定向三角剖分的定义[1]立刻推出，在所有的可定向三角剖分里必定存在异于零的闭链(曲面的情形是二维闭链)，而在不可定向的三角剖分里就没有这样的闭链．这个结果也可以立刻推广到任何维数的流形．

以上所引入的概念，使我们可以定义任意维数流形的一维、二维、…等等的连通阶．在每个流形的任意一个三角剖分里，同调独立的一维、二维、…闭链的最多个数，与该三角剖分的选择无关，叫作流形的一维、二维、…连通阶，或叫作一维、二维、…伯蒂数．

亏格为 p 的可定向闭曲面的一维伯蒂数等于 $2p$ (也就是如同在§2里所定义的连通阶)．射影平面的一维伯蒂数等于零(在射影平面里，任何不伸张到无穷远的闭链必包围一块面积，也就是同调于零．而伸张到无穷远的闭链，例如射影直线，当取它二次即二倍时同调于零)．一切不可定向曲面的二维伯蒂数等于零(在这种曲面上没有任何异于零的二维闭链)．

一切可定向二维曲面的伯蒂数等于1．事实上，如果把可定向曲面的某个三角剖分里的每一个三角形予以适当的定向，则得到一个闭链(所谓曲面的基本闭链)．不难验证，一切二维闭链都可以从把基本闭链乘上一个整数系数而得到．这个结果立刻可以推广到 n 维流形上去．还可以注意的是连通流形(也就是不分解为一块以上的流形)的零维伯蒂数算作是1．

各个维数的伯蒂数与欧拉特征数之间存在着一个奇妙的公式，这是欧拉定理的推广，为庞加莱所证明．这个著称的欧拉-庞加莱公式具有下列的简单形状：

$$\sum_{r=0}^{n}(-1)^r\alpha_r=\sum_{r=0}^{n}(-1)^r p_r.$$

---

1) 在前面，这个定义是对于曲面的三角剖分而给的，但立刻可以推广到任意维数流形的三角剖分上．

这里，左方是流形任意一个三角剖分的欧拉特征数，右方的数p_r是这个流形的r维伯蒂数. 特别，对于可定向曲面，我们方才已经看到$p_0=p_2=1$, $p_1=2p$, 而p是曲面的亏格. 从这里就得到关于可定向曲面的欧拉定理：

$$\alpha_0-\alpha_1+\alpha_2=2-2p.$$

§5. 向 量 场

考虑在某个平面区域G里给出的最简单的微分方程

$$\frac{dy}{dx}=F(x,\ y), \tag{2}$$

它的几何涵义是在区域G的每一点$(x,\ y)$给定了一个方向，这个方向的角系数等于$F(x,\ y)$. 这里$F(x,\ y)$是点$(x,\ y)$的连续函数，就是通常所谓在区域G里给定了一个连续的方向场. 我们可以很容易地把它变为一个向量场，例如，只要沿着每个以上所给定的方向取一个单位长度的向量. 微分方程(2)的积分问题实际上就是，如果这是可能的，把已给的平面区域分解为两两互不相交的曲线(方程的“积分曲线”). 使得在区域的每一点，切于在这一点给定的方向有唯一的一条积分曲线通过.

例如，考虑方程

$$\frac{dy}{dx}=\frac{y}{x}.$$

它在平面上每一点$M(x,y)$所确定的方向显然是射线$\overrightarrow{OM}$的方向(这里，O是坐标原点). 积分曲线是通过点O的直线. 通过平面上异于O的点有唯一的积分曲线. 至于坐标原点，则是这个微分方程的一个奇点(所谓“结点”)，所有的积分曲线都通过这一点.

如果我们取微分方程

$$\frac{dy}{dx}=-\frac{x}{y},$$

则可以看出，它在每个异于O的点$M(x,\ y)$所确定的方向垂直于$\overrightarrow{OM}$. 在这个情形，积分曲线是以点O为中心的圆，点O仍然是微

分方程的奇点，但是属于另外一种类型，它不是“结点”，而是所谓“中心”．还存在其他类型的奇点(见第五章，§6)，图 21，22 上画出了一些．微分方程 $\frac{dy}{dx}=\frac{y}{x}$ 没有闭积分曲线，但微分方程 $\frac{dy}{dx}=-\frac{x}{y}$ 只有闭积分曲线．也有那种积分曲线，螺旋式地绕紧奇点，这种奇点叫作焦点．

对于很多实用部门来说，极其重要的情形是所谓极限圈，这是一条闭积分曲线，而其他的积分曲线螺旋式地围着它逐渐绕紧．积分曲线相互之间相对位置的分布情况，积分曲线与奇点之间相对位置的分布情况，都还有许多其他的情形存在．凡是涉及微分方程积分曲线的形状与分布以及奇点数目、性质、相互分布位置等的问题，都是属于微分方程定性理论的．如同它的名称所表现的，微分方程的定性理论把直接对微分方程施行积分而得出“有限形式”解案的企图权且撇开一旁，正如有些用近似方法求数值积分的情况．定性理论的基本对象，实质上是微分方程的方向场以及积分曲线系的拓扑学．

首先是由于力学、物理与技术科学中问题的刺激才使得人们以定性的方式来接近微分方程，这中间包含闭积分曲线的存在问题，特别是一切有关极限圈的存在、数目与产生的问题．最初，这些问题的产生是由于庞加莱关于天体力学与宇宙学的研究．如同前面曾经说过，这些也正是促使这位法国几何学者从事拓扑学研究的动机．在苏联学者所创始的、属于振动论与无线电技术的卓越的研究工作中，微分方程论的拓扑问题占据了中心的位置之一．我所指的是 Л. И. 曼德尔斯塔姆的学派以及由那里传下来的 A. A. 安德洛诺夫学派，它们各自为成为微分方程定性理论的研究中心，并且可以说也是微分方程拓扑理论研究工作的主要中心之一．另一个在相当大的程度上用拓扑方法来研究微分方程定性理论的工作中心是莫斯科大学的 B. B. 斯捷巴诺夫与 B. B. 涅梅茨基学派．在列宁格勒，斯维尔德洛夫以及卡桑的数学家们关于微分方程定性理论的工作当中，拓扑方法也占有一个突出的地位．

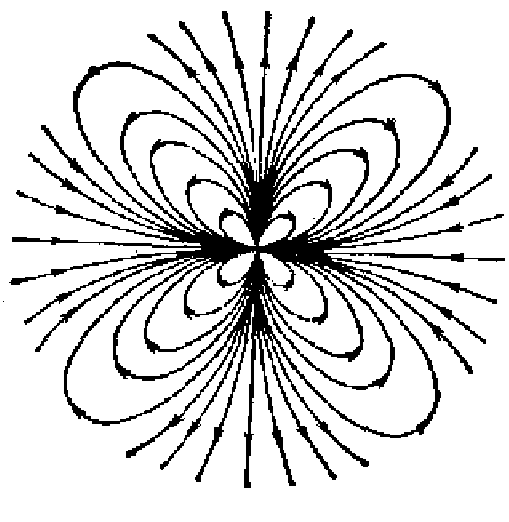


图 21

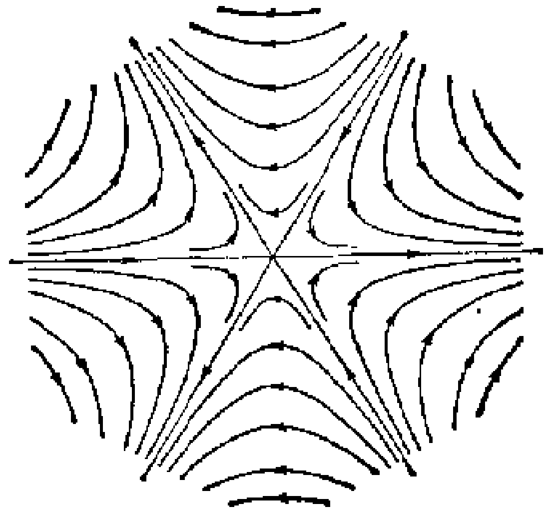


图 22

微分方程论不仅导致了平面上向量场的研究，并且也导致了高维流形上向量场的研究. 由多个微分方程所构成的最简单的方程组就已经几何地解释为高维欧氏空间里的方向场. 引入所谓方程的初积分，就意味着在所有的积分曲线中划分出那些位于某个流形上的曲线来，而流形是由初积分决定的. 任何动力体系(在它的古典意义之下)，一般地说，都导致了一个由它所有可能的状态全体所组成的高维流形(见§3)，并且导致了微分方程组，其积分曲线填满了由动力系统一切可能运动状态全体所形成的相空间，每一个单独的运动由这样或那样的初始条件决定. 因此，在这个情形，基本的研究对象也是流形上的方向场及轨线系. 大量的实际应用，特别是近几年来出现的，使得微分方程的定性理论在极其广泛的方面得到显著的发展. 因此，作为这个理论的基础的拓扑学也是如此. 近代拓扑学的迅速成长，以致过去半个世纪里在整个数学发展中占颇大的比重，也正应当归功于这些力学的、物理学的与天文学的课题所产生的推动力. 读者如果愿意知道微分方程论中拓扑问题在物理与技术方面的具体情况，可以参照 A. A. 安德洛诺夫与 C. Э. 海金合写的名著“振动论”.

作为在高维流形上向量场理论中所解决的具体问题的一个例子，考虑关于这种向量场奇点的代数总数问题.

设已给任意一个平滑流形. 为简单起见, 就假设是一个平滑的闭曲面. 假定在流形的每一点已给定了一个与它相切的向量, 并且无论按长度或按方向都随着该点连续地变动. 向量场的奇点就是流形上那样的点, 对于它们所给定的向量是零向量, 也就是说, 没有一个确定的方向. 我们假定每个这样的奇点是孤立点, 在闭流形上, 这就表示说只有有限多个奇点(否则, 这种点的一部分将聚在某个极限点周围, 由于连续性, 这极限点本身也是奇点, 但却不是孤立点).

对于孤立奇点可以定义它的指数——这个概念在某种意义之下类似于代数方程根的重数. 为了说明指数的定义, 把所考虑的奇点用某个闭曲线 C (在平面上就简单地是一个小圆) "孤立起来", 也就是说, 流形上这样的一条闭曲线, 它不通过任何奇点, 但包围了我们所考虑的那个奇点作为唯一的一个奇点在内部. 在曲线 C 上的每一点, 场的方向是唯一确定的. 为简单起见, 假设我们这个奇点的被 C 所围的邻域是一个平面区域(在一般的情形, 我们所留意的邻域以及它之上的场可以映在平面上). 若沿着曲线按正方向而环行一周, 则场的向量与某固定方向的交角比开始时增加的值必可写为 $2k\pi$ 的形状, 这里 k 是一个完全确定的整数. 这个整数就叫作奇点的指数, 也叫作向量沿着曲线 C 的旋度. 注意, 这个整数值不依赖于把奇点孤立起来所用的闭曲线的选择. 在图 21、22 上所画奇点的指数分别为 $+3$ 与 -2. 类似地, 但比较复杂一些, 可以定义在任意 n 维 ($n>2$) 流形上向量(方向)场奇点的指数. 德国数学家 H. 霍甫在 1926 年证明了下列美妙的定理: 如果在流形上所给的是只有有限多个奇点的连续向量场, 则这些奇点的指数之和, 或者所谓奇点的代数总数, 不依赖于向量场的选择, 并且总是等于流形的欧拉特征数.

从霍甫定理可以推知, 仅在欧拉特征数为零的流形上才可能给出不具有奇点的向量场; 并且可以证明, 在这样的流形上一定也可以造出不具有奇点的向量场. 因此, 在所有的闭曲面当中, 只有在环面以及所谓单侧环面(克莱茵曲面)上存在不具有奇点的向量

场*.

与向量场理论密切联系的是流形自己映入自己的连续映象的理论，特别是关于这种映象的不动点存在与否的问题．点 x 叫作映象 f 的不动点，假如它在映象 f 之下的象与它本身重合，也就是说

$$f(x)=x.$$

为要把这种联系说清楚，考虑一个最简单的情形——把圆 K 用连续映象 f 映入自己的情形．把圆内的每一点 x 与它的象 $f(x)$ 联结，我们得到一个向量 $\boldsymbol{u}_x=xf(x)$．这个向量为零，当且仅当 $f(x)=x$．也就是说，x 是映象 f 的不动点．我们来证明这样的点的确存在．设若不然，让我们来求沿着圆 K 的圆周 C 上述向量场的旋度．

当把向量场连续地变动时，显然它在圆周 C 上的旋度也连续地变动．旋度既然是一个整数，所以它也只能保持为常数．从这里就已经可以推出向量场沿着圆周 C 的旋度等于 1．事实上，因为圆 K 内每一点的象依然属于 K，所以对于圆周 C 上的点 x 来说，向量 $\boldsymbol{u}_x$(由于我们的假设，必不为零)指向圆内部，因此与半径 Ox 交于锐角，而半径 Ox 可以看作在点 x 的、指向圆心 O 的向量．

对于在圆周 C 上的一切点 x，令 $\boldsymbol{u}_x$ 受到这样的连续变动，使得它们每一个都逐渐地转动了一个锐角而达到最后与半径所代表的向量重合．我们方才讲过，经历这样的变动时，向量场沿着圆周 C 的旋度不变，但变动的结果，使得原来的向量场变为由半径组成的向量场，而后者的旋度显然是 1．因此，我们原来的向量场沿着圆周 C 的旋度也是 1．

由于所考虑的向量场是连续的，所以它在两个以 O 为中心而半径相差很小的同心圆上所得到的旋度值相等[1]．因此，这个向量场沿着圆 K 内一切以 O 为中心的圆周所取的旋度都具有相同的

---

* 克莱茵曲面上的向量场正负方向不分，因为它是无向曲面的一种．—校者注

1) 由假设映象 f 没有不动点，因此向量场到处有定义且异于零．这就有可能来说，向量场沿着圆 K 内任意一条闭曲线的旋度．

值1. 但因为由假设，向量 u_a 在圆内每一点有定义并且异于零，特别在原点 O 也是如此，因此，沿着以 O 为中心而且半径适当小的圆周，向量场所取的旋度值为零. 得到矛盾！于是推知，把圆映入自己的连续映象必定至少有一个不动点，这个定理是那个非常重要的布劳埃耳不动点定理的一个特别情形. 布劳埃耳定理说，任意一个把 n 维球体映入自己的连续映象必定至少有一个不动点.

这样或那样类型的关于不动点存在的定理，近年来已有详细的研讨，它们组成流形拓扑学的一个主要部分.

§6. 拓扑学的发展

在上个世纪末，拓扑学中或多或少得到一些发展的唯一分支是闭曲面的拓扑学. 这个理论的建立与十九世纪里复变函数论的发展相联系. 作为在十九世纪的数学史上最突出成绩之一的复变函数论，曾经循着好几条不同的道路而发展起来. 黎曼的几何方法，从掌握所研究对象的本质这一点来看，是最有成效的方法之一. 确凿地说明了在复变函数的一般理论里不能只限于考虑单值函数之后，黎曼方法引入了所谓黎曼曲面. 对于最简单的情形，象复变数的代数函数，则相应的这种曲面必然是可定向闭曲面. 在某种熟知的意义之下，研究这种曲面的拓扑性质就等价于研究所给的代数函数. 黎曼思想的进一步发展归功于庞加莱、克莱茵以及他们的后继者. 结果建立了函数论、闭曲面的拓扑学以及非欧几何（罗巴切夫斯基平面上运动群的理论）之间意想不到的深刻联系[1]. 于是，属于数学中完全不同分支的重大问题，第一次把拓扑学有机地包含进去.

当这些问题进一步发展后，单单考虑曲面的拓扑学便感到不足，而有必要解决 n 维拓扑学中一定的问题. 第一个遇到的这种问题是空间维数的拓扑不变性. 问题在于证明当 $n\neq m$ 时，n 维欧

1) 参照 A. И. 马库西维奨"解析函数论" 1950.

氏空间不能够拓扑地映满m维欧氏空间．这个困难的问题首先为布劳埃耳在1911年予以解决[1]．因为要解决这个问题而发明了一些新的拓扑方法，这些方法使得高维流形的连续映象理论以及高维流形上的向量场理论迅速地开始发展．在所有这些研究工作里，需要用到所谓点集拓扑学的最基本概念．拓扑学的这一分支，是由上世纪末叶康托尔所建立的集合论而产生的点集拓扑学所研究的对象，也就是说，所考虑的几何图形类是极其广泛的，即便不包括位于欧氏空间里的一切集合，至少也包括了欧氏空间中的一切闭集．对于新的点集拓扑学的迅速发展，各国学者都起了作用，特别应该指出的是波兰的拓扑学派．

在苏维埃拓扑学家的工作里，点集拓扑学的发展得到本质上全新的方向，特别是杰出的、但不幸夭折的苏联数学家 П. С. 乌利逊 (1898—1924) 所建立的一般维数论，使最一般的点集都能按基本的特征——维数来分类．这种分类极其有成效，并引起了以全新的一种观点来研究最广义的几何形状[2]．乌利逊在他的维数理论中所发展的思想，也成为使 Л. А. 刘斯铁尔尼克（与 Л. Г. 史尼雷尔曼合作的）在变分学方面卓越的工作所赖以产生的基础．

在这些工作里除了其他结果以外，还对庞加莱关于在一切同胚于球面的曲面上是否必然存在三条无重点闭测地线的问题给予了彻底的解决．

另一方面，在维数理论的基础上，П. С. 亚历山大洛夫把组合拓扑学里的代数方法移植到点集论的领域里来，而后者又引出了拓扑学研究的新方向．在这方面，直到最年轻一代的数学家，苏联

---

1) 实际上，为了复变函数论的发展还必须解决一个更难些的问题，那就是证明n维区域在n维空间里的拓扑象必然是一个区域．这个问题也为布劳埃耳解决了．

2) 乌利逊对于维数以归纳的方式所下的定义可以看作是罗巴切夫斯基把截断作为基本几何运算这种想法的一个最完全的发展．这个定义可以比较近似地陈述如下：集合是0维的，假如它可以表示为两两互不附贴的适当小子集之和，集合是n维的，假如它可以用$(n-1)$维子集"切"为适当小的两两互不附贴的子集之和，并且用小于$(n-1)$维的集合是不可能作到的（在§7里将给出，从近代拓扑学眼光来看较为精确的关于附贴性的定义）．

都始终领先[1].

至于组合拓扑学本身，大约在1915年前后，美国的拓扑学者魏布伦、比尔可夫、亚历山大、莱夫舍茨等人开始了一系列的研究.例如，亚历山大证明了伯蒂数的拓扑不变性以及他自己的基本对偶定理，后者是Л. С. 邦特里雅金进一步深入研究的出发点；莱夫舍茨给出了关于流形上任意自我连续映象的不动点代数总数的著名公式，并建立了连续映象一般代数理论的基础，后来又为霍甫进一步发展；比尔可夫在科学上的功绩是把动力系统的理论在它的拓扑方面与度量方面作了本质的推进，等等.流形拓扑学以及连续映象的理论在霍甫的工作里得到深远的发展，和一些别的结果一样，他证明存在无穷多个实质上互异的从三维球面到二维球面的连续映象. 所谓两个连续映象实质上互异，是指不能经过连续地变动把它们中的一个变为另一个. 这样，霍甫就奠定了一个新方向的基础，就是所谓同伦论. 近年来，在同伦论里，或者更一般地说，在整个组合拓扑学里，由新兴法国拓扑学派(莱雷，谢尔)工作的影响而获得了巨大的跃进.

如同前面所指出，乌利逊的基本的研究工作是苏维埃数学家在拓扑学领域里从事活动的开端. 这些研究属于点集拓扑学的范围，不过在二十年代末，苏维埃拓扑学家们已经把组合拓扑学也包含在自己的兴趣范围之内. 这个过程的进行是非常独特的，是由于把组合方法用来研究闭集，也就是说，由非常广义的对象而发生的. 在这个基础上产生了本世纪最突出的几何发现之一——Л. С. 邦特里雅金一般对偶律的陈述与证明. n维欧氏空间中闭集的拓扑结构与它余集的拓扑结构之间的关系可以说已经详尽无遗地包含在这个深刻的对偶律里了. 与他的对偶律相关联，Л. С. 邦特里雅金建立了交换群的一般特征标理论. 这又引导他对一般拓扑群

1) 这里应该指出П. С. 亚历山大洛夫的所谓维数的同调理论，与之有关的为Л. С. 邦特里雅金所构出的有趣的例子，以及在M. Ф. 波克斯坦、B. Г. 保尔强斯基，特别是K. A. 西特尼可夫的工作中关于维数的同调理论进一步的发展. Л. С. 邦特里雅金的对偶律，我们即将谈到.

与古典的连续李群进行研究. 这个领域由于 Л. C. 邦特里雅金的工作而完全改变了面貌. 此外, Л. C. 邦特里雅金与他的学生 (B. Г. 保尔强斯基, M. M. 波斯特尼可夫等人)对于流形的拓扑学以及它们的连续映象作了许多卓越的工作. 在这些工作里, 一种所谓 ▽ 同调的新方法得到应用. ▽ 同调是 A. H. 阔尔莫果洛夫与亚历山大分别独立地引入组合拓扑学的. 这个现在在整个同调论里占头等地位的方法, 也使得 Л. C. 邦特里雅金对偶律得以在完全不同的方向继续发展, 因而引出 A. H. 阔尔莫果洛夫(与亚历山大)的、П. C. 亚历山大洛夫的以及 K. A. 西特尼可夫的对偶律, 这些都是近代拓扑学中突出的结果. 在 Л. A. 刘斯铁尔尼克关于变分学的最近工作里, 这个方法也得到重要的应用.

§7. 度量空间与拓扑空间

在这一章的开始, 我们曾经谈到把(图形各部分的)附贴关系作为基本的拓扑概念, 把连续映象定义作保持附贴关系的变换, 但是我们并没有给出这个基本概念的精确定义. 若要把这个定义给得足够广泛, 也只有依赖集合论的概念. 本节的目的正在于此, 引入拓扑空间的概念就使这个问题圆满解决.

集合论所能够给出的关于几何图形的概念, 其广泛的程度与一般性是所谓"古典"数学所达不到的. 几何学的, 特别是拓扑学的研究对象现在变成了任意的点集, 也就是说, 任意一个以 n 维欧氏空间中的点作为元素的集合. n 维空间的点与点之间有确定的距离, 点 $A=(x_1, x_2, \cdots, x_n)$ 与点 $B=(y_1, y_2, \cdots, y_n)$ 之间的距离按定义等于非负实数

$$\rho(A, B)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n+y_n)^2}.$$

距离概念使得我们可以第一步先定义集合与点的附贴, 然后定义两个集合之间的附贴. 我们说点 A 是集合 M 的附贴点, 假如在集合 M 里存在与点 A 的距离小于任意预先给定正数的点. 显然, 集合中的每一点是附贴点, 但不属于集合的点也可以是附贴

点．例如，取数轴上的开区间(0, 1)，也就是说，位于0与1之间的一切点；点0与1本身并不属于这个区间，但在区间(0, 1)中存在着任意地接近于0与1的点．如果一个集合包含了它所有的附贴点，则这个集合叫作闭集．例如，数轴上的区间[0, 1]，所有满足不等式$0\leqslant x\leqslant 1$的点，组成一个闭集．

平面上的或者三维空间里的或者更高维空间里的闭集可以是构造极其复杂的集合．正是这种集合在n维空间的点集拓扑学里将被作为首要的研究对象．我们说两个集合P与Q互相附贴，假如它们之中至少有一个包含了另一个的附贴点．根据这个定义，两个相交的(具有公共点的)集合总是互相附贴的．从上面的定义可知，两个闭集仅当具有公共点时才附贴；又如区间[0, 1]与(1, 2)则是附贴的，虽然它们不具有公共点．但区间[0, 1]中的点1同时也是区间(1, 2)中的附贴点．其次，我们说集合R被它的子集S分割(“截断”)，或者说S是集合R的“截面”，假如R中所有的不属于S的点全体所形成的集合$R-S$可以分解为两个互不附贴集合的和．

于是，罗巴切夫斯基关于附贴与截断的想法就在近代拓扑学里得到高度精确与广泛的表达．我已经看到，乌利逊怎样以这些想法为基础而定义了任意集合的维数(见211页注释)．现在这个定义也得到完全精确的涵义．对于连续映象或变换来说，情况也是如此．把集合X映入集合Y的映象f叫作是连续的，假如f保持附贴性，也就是说，如果集合X中的点A是X中某个子集p的附贴点，则点A的象$f(A)$是集合p的象$f(p)$的附贴点．最后把集合X一对一地映满集合Y的映象叫作是拓扑的，假如这个映象自己以及它的逆映象(把Y映满X)都是连续映象．这些定义使得我们在第一节里所说的事情得到完全确切的基础．

但是点集拓扑学广泛的程度决不限于只考虑由通常意义的点所构成的几何图形．距离概念不仅仅可以在欧氏空间里的点与点之间引入，并且可以在其他的、看起来与几何毫不相干的那种对象上引进．

例如，考虑在区间 (0, 1) 上定义的连续函数全体所形成的集合，定义函数 f 与 g 之间的距离 $\rho(f, g)$ 为式子 $|f(x)-g(x)|$ 当 x 遍及整个区间(0,1)时所取的最大值. 这个“距离”满足空间中两点之间距离的一切基本性质：两个函数 f 与 g 之间的距离 $\rho(f,g)$ 等于零，当且仅当这两个函数重合，也就是说，当 $f(x)=g(x)$ 对于一切 x 成立；其次，距离显然是对称的，也就是说，$\rho(f, g)=\rho(g, f)$；最后，它满足所谓三角形公理，即对于任意三个函数 f_1, f_2, f_3, 我们有 $\rho(f_1, f_2)+\rho(f_2, f_3)\geqslant\rho(f_1, f_3)$. 通常说，引进距离定义使我们的函数集合成为度量空间(一般记作 C). 所谓度量空间，一般了解如下. 由某种叫作点的对象所构成的集合，如果对于每两点定义了一个非负实数叫作**距离**，满足刚才所说的“距离公理”，则我们说这个集合具备了上述的距离以后，形成一个度量空间.

对于任何度量空间，我们可以考虑它子集的附贴点，考虑子集之间的附贴，因此一般地说，可以考虑拓扑概念(闭集，连续映象，以及由这些最简单的概念所演出的概念). 这种方式为拓扑的想法或一般地说几何的想法开辟了广阔而丰富的应用领域. 初看起来，好象这些领域与几何的想法完全无关，我们举例说明之.

仍然取微分方程(2)：

$$\frac{dy}{dx}=F(x, y).$$

如果对于 $0\leqslant x\leqslant 1$, $y=\varphi(x)$ 是方程的解，并且当 $x=0$ 时 $y=0$, 则显然函数 $\varphi(x)$ 满足积分方程

$$\varphi(x)=\int_0^x F(x, \varphi(x))dx. \tag{3}$$

现在考虑积分 $G(f)=\int_0^x F(x, f(x))dx$, 这里 $0\leqslant x\leqslant 1$, 而 $f(x)$ 是任意一个在闭区间(0, 1)上定义的连续函数. 这个积分也是在区间 (0, 1) 上定义的某一个连续函数 $g(x)$. 这样，式子 $G(f)=\int_0^x F(x, f(x))dx$ 对于每个函数 f 对应了一个函数 $g=G(f)$, 换句

话说，我们有一个映象 G 将度量空间 C 映入自己，并且容易看出来，这个映象是连续的，这时，作为方程(2)或与之等价的方程(3)的解，函数 $\varphi(x)$(可能有好几个)将有些什么特征性质呢？显然可以看出，在我们的映象 G 之下它变为自己，也就是说，是映象 G 的不动点．映象 G 的确存在这样的不动点——这可以从 1926 年 П. С. 亚历山大洛夫与 В. В. 涅梅茨基所证明关于度量空间自我连续映象不动点的一个非常一般的定理推出．近年来，以这种或那种函数为点的度量空间(这种空间叫作函数空间)已成为数学分析中经常应用的工具；以一部分拓扑方法，但主要是用广义的代数方法来研究函数空间正是泛函分析的内容(见第十九章)．

如引言里所指出的，泛函分析在近代数学里占有极其重要的地位，这是由于它与其他数学分支的联系很广泛，并且由于它对自然科学的价值，首先是对于理论物理、函数空间拓扑性质的研究与变分学、偏微分方程论有密切的联系 (Л. А. 刘斯铁尔尼克、莫尔氏、莱雷、晓德尔、М. А. 克拉斯诺塞尔斯基)．函数空间中自我连续映象的不动点存在问题在这些研究工作中很有地位．

函数空间与更一般些的度量空间的拓扑学，就其一般性的程度来说还没有达到近代拓扑理论的最后界限，问题在于度量空间中的基本拓扑概念——附贴性——是依靠点之间的距离而引进的，但距离本身并不是拓扑概念．因此就产生了直接用公理法来定义附贴性，这样就引出了拓扑空间的概念——近代拓扑学中最一般的空间概念．

拓扑空间是由任意属性的元素(叫作拓扑空间的点)所组成的一个集合，并对于它的每个子集以这种或那种方式给出了它的附贴点，这时，还必须添加一些自然的条件，叫作拓扑空间的公理(例如，集合的每一点是它的附贴点，两个集合和集的附贴点必为这两个集合中之一个的附贴点，等等)．近年来，拓扑空间的理论已经研究得很深入，它的发展，苏维埃数学家 П. С. 乌利逊、П. С. 亚历山大洛夫、А. Н. 吉洪诺夫等人起了首要的作用．在拓扑空间论最新的具有原则性意义的结果中，应该指出青年数学家 Ю. М. 斯

米尔诺夫所发现的拓扑空间可度量化的必要与充分条件；也就是说，在这种条件之下，拓扑空间的每两点之间可以定义距离，使得空间原来具备的“拓扑结构”可以看作是由这个距离概念所诱导的；换句话说，使得在所得到的度量空间里，任意子集的附贴点也正好是在原先给的拓扑空间里这个子集的附贴点。

文　　献

П. С. Александров, Комбинаторная топология. Гостехиздат, 1947.

这是一本有系统的、包罗较广的专著，对读者的水平有较高的要求。

П. С. Александров, Теоремы двойственности в комбинаторной топологии, т. Изд-во АН СССР, 1947.

伟大的十月社会主义革命三十周年纪念文集。

П. С. 亚历山大洛夫，集与函数的泛论初阶，高等教育出版社，1955.

这本书是点集拓扑学(度量与拓扑空间)的导论。

П. С. 亚历山大洛夫，拓扑对偶定理，科学出版社，1959.

可以作为学习组合拓扑学的第一本书，它把读者引向拓扑学一个重要分支(对偶定理)中最近的研究工作。

П. С. Александров и В. А. Ефремович, Очерк основных понятий топологии, М. Л., 1936.

这是一本介绍组合拓扑学的书，对于曲面拓扑学有较详尽的叙述。

Л. С. 邦特里雅金，组合拓扑学基础，科学出版社，1954.

对于 n 维组合拓扑学的基本内容，作了扼要而又完整无遗的叙述。

孙以丰　译

秦元勋　校

第十九章 泛函分析

泛函分析在二十世纪中的产生和发展，实质上受着两个因素的推动．一方面，出现了用统一的观点来理解十九世纪数学各个分支(常是很少相互联系)所积累的大量实际材料的必要性．由此泛函分析的基本概念从不同的方面和不同的联系中产生了，而且这些概念随后就固定下来．泛函分析的许多基本概念自然地出现在变分学发展的进程中，出现在关于振动的问题(从具有穷个自由度体系的振动到连续介质的振动的过渡)中，出现在积分方程理论中，出现在常、偏微分方程理论(边界问题、关于固有值的问题等等)中，出现在实变数函数论的发展中，出现在算符法*中，出现在函数逼近论问题的研究中，等等．泛函分析使得人们能用统一的观点来理解在这些领域中可利用的许多结果，并且常常促成新结果的获得．由此建立起来的概念和方法，最近几十年来已被用于理论物理学的新分支——量子力学中．

另一方面，与量子力学有关的数学问题的研究，表现为泛函分析进一步发展中的转折点，它们已经形成且在当代继续形成泛函分析发展的基本方向．

泛函分析离它的完备还很远．然而无疑地，在它的进一步发展中，近代物理学的问题和要求将对于它有这样的意义，就像古典力学对于微积分在十八世纪中的产生和发展一样．

我们不想在这一章里介绍泛函分析的一切基本问题．许多重要的分支这一章还没有谈到．我们现在只就一些拣选过的问题，给读者介绍一些泛函分析的基本概念，并且尽可能地阐明前面提到的关于它们的联系．这些问题的实质二十世纪初在希尔伯特——泛函分析创始人之一——的经典著作中已有了根本性的讨

* 即运算微积．——译者注

论. 从那时起, 泛函分析有力地发展了, 并且广泛地被应用于几乎所有的数学分支中: 被应用于偏微分方程论中、概率论中、量子力学中、量子场论中等等. 很遗憾, 泛函分析这种进一步的发展没有能在这里反映出. 为了描述一下这种进一步的发展, 必须写出许多文章. 因此, 我们就限于讨论其中一个最古老的问题——固有函数理论.

§1. n 维 空 间

以下我们将要利用关于 n 维空间的基本概念. 虽然在线性代数及抽象空间的两章中已经引入了这些概念, 我们认为, 用今后将要遇到的形式在这里简短地重复一下这些概念还是有用的. 为了阅读这一段, 读者只要知道解析几何概要就够了.

我们知道, 在三维空间几何中, 点是由三个数 (f_1, f_2, f_3)——点的坐标——给出的. 由这个点到坐标原点的距离等于 $\sqrt{f_1^2+f_2^2+f_3^2}$. 假如用向量的端点来表示点, 向量是由坐标原点引向这个点, 那么向量的长也等于 $\sqrt{f_1^2+f_2^2+f_3^2}$. 由坐标原点引到两个不同的点 $A(f_1, f_2, f_3)$ 及 $B(g_1, g_2, g_3)$ 的两个向量间夹角的余弦由以下公式决定:

$$\cos\varphi=\frac{f_1g_1+f_2g_2+f_3g_3}{\sqrt{f_1^2+f_2^2+f_3^2}\sqrt{g_1^2+g_2^2+g_3^2}}.$$

从三角学我们知道 $|\cos\varphi|\leqslant 1$. 因此有不等式

$$\frac{|f_1g_1+f_2g_2+f_3g_3|}{\sqrt{f_1^2+f_2^2+f_3^2}\sqrt{g_1^2+g_2^2+g_3^2}}\leqslant 1,$$

意即永远有

$$(f_1g_1+f_2g_2+f_3g_3)^2\leqslant(f_1^2+f_2^2+f_3^2)\cdot(g_1^2+g_2^2+g_3^2). \qquad (1)$$

最后这个不等式带有代数的特性, 并且对于任何六个数 (f_1, f_2, f_3) 及 (g_1, g_2, g_3) 都对, 因为任何六个数可以当做空间中两点的坐标. 不等式 (1) 是从纯几何学的考虑得来且与几何学密切相关的, 这使人能把它了解得更清楚.

在一系列几何学的关系的解析提法中，常常可以看出来：当用 n 个数来代替三个数时相应的事实也对．举例说，前面引入的不等式 (1) 可以推广到 $2n$ 个数 $(f_1, f_2, \cdots, f_n)$ 及 $(g_1, g_2, \cdots, g_n)$ 的场合．这意思是对于任何 $2n$ 个数 $(f_1, f_2, \cdots, f_n)$ 及 $(g_1, g_2, \cdots, g_n)$，与 (1) 类似的不等式也对，就是

$$(f_1g_1+f_2g_2+\cdots+f_ng_n)^2\leqslant(f_1^2+f_2^2+\cdots+f_n^2)(g_1^2+g_2^2+\cdots+g_n^2). \tag{1'}$$

这不等式，以 (1) 为其特殊情形，可纯解析地证明[1]．类似地，把三个数与三个数间许多其它关系推广到 n 个数上面，数与数间的关系（数量关系），与几何学的这种联系，当引用 n 维空间概念时，就显得特别突出．这个概念已在第十六章中引进．我们在这里扼要地将它重复一下．

n 个数的总体 $(f_1, f_2, \cdots, f_n)$ 称为 n 维空间的点或向量（我们将常使用后一叫法）．向量 $(f_1, f_2, \cdots, f_n)$ 将在以后简短地用一个字母 f 来表示．

像在三维空间里向量的加法一样，向量

$$f=\{f_1, f_2, \cdots, f_n\} \text{ 及 } g=\{g_1, g_2, \cdots, g_n\}$$

之和定义为向量 $\{f_1+g_1, f_2+g_2, \cdots, f_n+g_n\}$，且用 $f+g$ 来表示．

向量 $f=\{f_1, f_2, \cdots, f_n\}$ 与数 λ 的乘积定义为向量

$$\lambda f=\{\lambda f_1, \lambda f_2, \cdots, \lambda f_n\}.$$

向量 $f=\{f_1, f_2, \cdots, f_n\}$ 的长与三维空间中向量的长类似，定义为 $\sqrt{f_1^2+f_2^2+\cdots+f_n^2}$．

n 维空间中两个向量 $f=\{f_1, f_2, \cdots, f_n\}$ 及 $g=\{g_1, g_2, \cdots, g_n\}$ 间的角度 φ 由它的余弦给出，也完全像在三维空间中向量间的角度那样定义．就是说，它由下列公式定义：

$$\cos\varphi=\frac{f_1g_1+f_2g_2+\cdots+f_ng_n}{\sqrt{f_1^2+f_2^2+\cdots+f_n^2}\sqrt{g_1^2+g_2^2+\cdots+g_n^2}}\,[2]. \tag{2}$$

两向量的纯积是指这两个向量的长与其间角度余弦的乘积．

---

1) 参阅本书第 55 页．

2) $|\cos\varphi|\leqslant1$ 由不等式 (1′) 推出．

依照这个方式，假如 $f=\{f_1, f_2, \cdots, f_n\}$ 及 $g=\{g_1, g_2, \cdots, g_n\}$，则因向量的长分别等于

$$\sqrt{f_1^2+f_2^2+\cdots+f_n^2} \text{ 及 } \sqrt{g_1^2+g_2^2+\cdots+g_n^2},$$

所以，它们的纯积 (f, g) 由下列公式给出：

$$(f, g)=f_1g_1+f_2g_2+\cdots+f_ng_n. \tag{3}$$

特别地，两向量直交性表现为等式 $\cos\varphi=0$，即 $(f, g)=0$.

借助于公式 (3)，读者可以看出，在 n 维空间中纯积具有以下性质：

1. $(f, g)=(g, f)$.

2. $(\lambda f, g)=\lambda(f, g)$.

3. $(f, g_1+g_2)=(f, g_1)+(f, g_2)$.

4. $(f, f)\geqslant 0$，此处等号仅当 $f=0$，即 $f_1=f_2=\cdots=f_n=0$ 时成立.

向量 f 与其自身的纯积 (f, f) 等于向量 f 的长的平方.

纯积的 n 维空间的研究中是很有用的工具. 我们在这里将不一般地研究 n 维空间的几何，而仅限于研究一个例子.

作为这样的例子，考虑 n 维空间中的毕达哥拉斯定理：斜边的平方等于直边的平方和. 为了证明这一定理，我们选取在平面上这个定理的这样一种证明方式，使得这种证明方式易于被搬到 n 维空间的情形中.

设 f 及 g 是平面上两个互相垂直的向量. 考虑在向量 f 及 g 上构成的直角三角形(图 1). 这个三角形的斜边的长等于向量

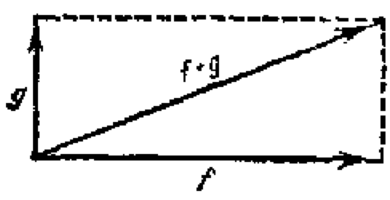


图 1

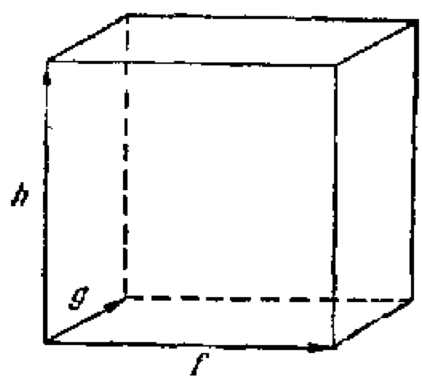


图 2

$f+g$ 的长．依我们的表记法用向量方式来写毕达哥拉斯定理．因为向量长的平方等于向量与其自身的纯积，所以用纯积这名词可将毕达哥拉斯定理写做

$$(f+g,\ f+g)=(f,\ f)+(g,\ g).$$

从纯积的性质立刻得到证明．事实上，

$$(f+g,\ f+g)=(f,\ f)+(f,\ g)+(g,\ f)+(g,\ g).$$

而右端中间两项由于向量 f 与 g 的直交性等于 0.

在上面所引的证明中，我们仅利用了向量的长、向量的直交性定义和纯积的性质．因之，如果我们假定 f 及 g 是 n 维空间中的两直交向量，证明中没有什么需要改变的地方．就这样毕达哥拉斯定理对于 n 维空间的直角三角形得到了证明．

假如在 n 维空间中给了两两直交的向量 $f,\ g$ 及 h，则这些向量的和 $f+g+h$ 代表在这些向量上形成的长方体的对角线(图2)，并且有等式

$$(f+g+h,\ f+g+h)=(f,\ f)+(g,\ g)+(h,\ h).$$

它的意思就是：长方体对角线的长的平方等于它的棱边的平方和．这个论断的证明完全类似于上面所引的毕达哥拉斯定理的证明，我们把它留给读者．完全一样，假如在 n 维空间中有 k 个两两直交的向量 $f^1,\ f^2,\ \cdots,\ f^k$，则一样简单地可证明等式

$$\begin{aligned}&(f^1+f^2+\cdots+f^k,\ f^1+f^2+\cdots+f^k)\\&\quad=(f^1,\ f^1)+(f^2,\ f^2)+\cdots+(f^k,\ f^k),\end{aligned}\tag{4}$$

它的意思是说：在 n 维空间中，"k-维长方体"的对角线平方也等于它的棱边平方和．

§2．希尔伯特空间(无穷维空间)

与 n 维空间的联系　n 维空间的概念，在数学和物理学的一系列问题的研究中是很有用的．转过来，这个概念又推动空间概念的进一步发展和它在数学各个部门中的应用．在线性代数和 n 维空间几何学的发展中，关于弹性系统微小振动的问题起过很大



图 3

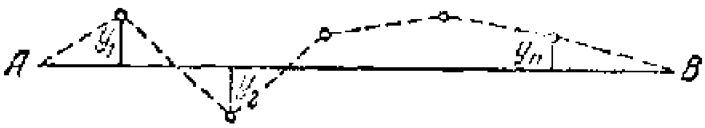


图 4

的作用.

考虑这种问题的下述典型例子(图 3). 设 AB 为柔韧的线, 在点 A 及 B 之间拉紧. 设想在线的某点 C 处系一个重荷. 假如把它从平衡的位置放开, 它就开始以频率 ω 振动. 当已知线的张力、质量 m 和线的位置时, 就可以计算频率的大小. 在每一时刻, 系统的位置由一个数决定, 即由从质量 m 的位置到线的平衡位置的离差 y_1 决定.

在线 AB 上配置 n 个重荷于点 $C_1, C_2, \cdots, C_n$. 在此情形中我们认为线无重量. 这意思是说, 它的质量是这样小, 与所配置的重荷的质量相比较, 我们可以将它忽略. 这个系统的位置由 n 个数 $y_1, y_2, \cdots, y_n$(即所系重荷到平衡位置的离差)给出. 数 $y_1, y_2, \cdots, y_n$ 的总体可看做 n 维空间中的向量(这在许多关系中都是有用的).

上例中引入的微小振动的研究, 恰当地显现出与 n 维空间几何学基本事实的联系. 例如, 确定这一系统的振动频率的问题可化归为确定 n 维空间中某一椭球体的轴的位置问题.

现在考虑在点 A 及 B 间拉紧的弦的微小振动问题. 在此情形中, 我们把理想化的弦看作一条柔韧的线, 即质量为有穷的连续地沿线分布的线. 特别是, 在均匀的弦这名称下, 我们理解为具定常密度的弦.

因为质量是沿着弦连续分布的, 所以弦的位置不能由有穷多个数 $y_1, y_2, \cdots, y_n$ 给出, 而须给出弦的每点 x 的离差 $y(x)$. 依此方式, 在每个时刻弦的位置由某一函数 $y(x)$ 给出.

在坐标为 $x_1, x_2, \cdots, x_n$ 的点处有 n 个系重的线的位置, 由具 n 个连结点的折线图象描绘(图 4). 假如系重的数目增大, 则相应地折线的连结点数增大. 假如系重的数目无穷增大, 而且相邻

接的重荷间距离趋于 0，则在极限情形我们得到的就是质量沿着线连续分布的情形，也就是理想化的弦的情形. 描绘带重荷的线的位置的折线，在此情形中化为描绘弦的位置的曲线(图 5).

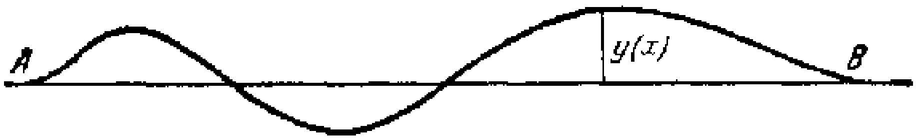


图 5

像这样，我们看到，在带重荷的线的振动与弦的振动间存在着密切的联系. 在第一种问题中，系统的位置由点或 n 维空间中的向量给出. 因之，在第二种问题中，很自然地，给出弦的位置的函数也可以看成某一无穷维空间中的向量或点. 一系列的类似问题[1] 化归到这样地考虑空间的观点上，而空间的点（向量）则代表定义在某区间上的函数.

以上所考虑的关于振动问题的例（在 § 4 中我们还要回转来考虑它）向我们指明，无穷维空间中的基本概念应当怎样引入.

希尔伯特空间 在这里我们考虑应用最广且最重要的无穷维空间概念之一，就是希尔伯特空间.

n 维空间中的向量定义为 n 个数 f_i 的总体，此处 i 从 1 变到 n. 类似地，无穷维空间中的向量定义为函数 $f(x)$，此处 x 从 a

1) 作为一个这样的问题的例，考虑由相互链式连接的电流回路产生的电振动(图 6).

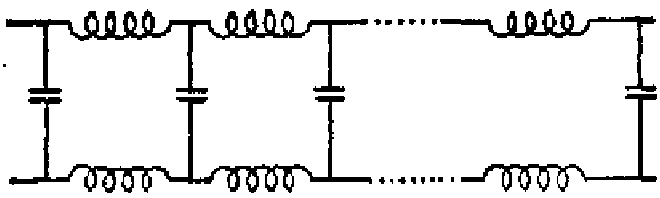


图 6

这样的电路状态由 n 个数 $u_1, u_2, \cdots, u_n$ 表示，此处 u_i 是在第 i 个回路的蓄电器上的电动势. n 个数的总体 $(u_1, u_2, \cdots, u_n)$ 可看作 n 维空间中的向量.

设有一个双线制线路，即由两个导线组成的，具有沿着导线分布的一定的最终电容及电感. 这种线路的电状态由一个函数 $u(x)$ 表现出来，$u(x)$ 给出沿线分布的电压. 这个函数是定义在某区间 (a, b) 上的函数的无穷维空间的向量.

变到 b.

向量的加法和数与向量的乘法定义为函数的加法和函数与数的乘法.

n 维空间中的向量 f 的长由公式 $\sqrt{\sum_{i=1}^{n} f_i^2}$ 定义. 因为对于函数说积分起着求和的作用，所以希尔伯特空间中的向量 $f(x)$ 的长由公式

$$\sqrt{\int_a^b f^2(x)dx} \tag{5}$$

给出.

在 n 维空间中点 f 及 g 间的距离定义为向量 $f-g$ 的长，即 $\sqrt{\sum_{i=1}^{n}(f_i-g_i)^2}$. 类似地，在泛函空间中元素 $f(t)$ 及 $g(t)$ 间的"距离"等于 $\sqrt{\int_a^b [f(t)-g(t)]^2 dt}$.

表达式 $\sqrt{\int_a^b [f(t)-g(t)]^2 dt}$ 称为函数 $f(t)$ 及 $g(t)$ 的均方差. 就像这样，以均方差作为希尔伯特空间中两元素间距离的量度.

转到两向量间角度的定义上来. 在 n 维空间中向量 $f=\{f_i\}$ 及 $g=\{g_i\}$ 间的角度 φ 由下列公式定义:

$$\cos\varphi=\frac{\sum_{i=1}^{n} f_i g_i}{\sqrt{\sum_{i=1}^{n} f_i^2}\sqrt{\sum_{i=1}^{n} g_i^2}}.$$

在无穷维空间中求和由相应的积分来代替，而希尔伯特空间中两向量 f 及 g 间的角度 φ 由类似的公式定义:

$$\cos\varphi=\frac{\int_a^b f(t)g(t)dt}{\sqrt{\int_a^b f^2(t)dt}\sqrt{\int_a^b g^2(t)dt}}. \tag{6}$$

假如右端的绝对值不大于 1，即假如

$$\left|\int_a^b f(t)g(t)dt\right| \leqslant \sqrt{\int_a^b f^2(t)dt}\sqrt{\int_a^b g^2(t)dt}, \tag{7}$$

则表达式(6)可看成某角度 φ 的余弦. 不等式(7)事实上对于任意

两个函数 $f(t)$ 及 $g(t)$ 成立. 它在分析中起重要作用，被称为柯西-布涅科夫斯基不等式. 现在我们来证明它.

设 $f(x)$ 及 $g(x)$ 是两个定义在 (a, b) 上不恒等于零的函数，取任意数 λ 及 μ 并组成表达式

$$\int_a^b [\lambda f(x) - \mu g(x)]^2 dx.$$

因为在积分号下函数 $[\lambda f(x) - \mu g(x)]^2$ 非负，故有下列不等式成立：

$$\int_a^b [\lambda f(x) - \mu g(x)]^2 dx \geqslant 0,$$

即
$$2\lambda\mu \int_a^b f(x) g(x) dx \leqslant \lambda^2 \int_a^b f^2(x) dx + \mu^2 \int_a^b g^2(x) dx.$$

为简短计引入表记法

$$\left|\int_a^b f(x) g(x) dx\right| = C, \quad \int_a^b f^2(x) dx = A,$$

$$\int_a^b g^2(x) dx = B. \tag{8}$$

用这些表记法，不等式可改写为下列形式[1]：

$$2\lambda\mu C \leqslant \lambda^2 A + \mu^2 B. \tag{9}$$

这不等式对于任何数值 λ 及 μ 保持真确，特别是可令

$$\lambda = \sqrt{\frac{C}{A}}, \quad \mu = \sqrt{\frac{C}{B}}. \tag{10}$$

把 λ 及 μ 的值代入不等式(9)得

$$\frac{C}{\sqrt{AB}} \leqslant 1.$$

按(8)式将 A, B, C 所代表的积分代入，最后得柯西-布涅科夫斯基不等式.

在几何学中向量的纯积定义为它们的长同它们之间角度的余弦的乘积. 向量 f 及 g 的长在我们的情形中等于

$$\sqrt{\int_a^b f^2(x) dx} \text{ 及 } \sqrt{\int_a^b g^2(x) dx},$$

1) 考虑到 λ 或 μ 的符号的任意性，我们可以取积分的绝对值作为 C.

而它们之间角度的余弦由下列公式定义：

$$\cos\varphi=\frac{\int_a^b f(x)g(x)dx}{\sqrt{\int_a^b f^2(x)dx}\sqrt{\int_a^b g^2(x)dx}}.$$

这些表式相乘，我们得到希尔伯特空间中的两向量的纯积公式：

$$(f,\ g)=\int_a^b f(x)g(x)dx. \tag{11}$$

从这公式看出，向量 f 与其自身的纯积是它的长度的平方。

假如非零向量 f 及 g 的纯积等于 0，那末它的意义是 $\cos\varphi=0$，就是用我们的定义来说，它们所给的角度 φ 等于 90°。因之向量 f 及 g 称为直交的，这时

$$(f,\ g)=\int_a^b f(x)g(x)dx=0.$$

在希尔伯特空间中，像在 n 维空间中一样，毕达哥拉斯定理真确(参见 § 1)。设 $f_1(x)$，$f_2(x)$，…，$f_N(x)$ 表两两直交的函数，而

$$f(x)=f_1(x)+f_2(x)+\cdots+f_N(x)$$

为它们的和，则 f 的长度平方等于 f_1，f_2，…，f_N 的长度平方和。

因为在希尔伯特空间中向量的长借助于积分而给出，故在这种情形中毕达哥拉斯定理由下列公式表出：

$$\int_a^b f^2(x)dx=\int_a^b f_1^2(x)dx+\int_a^b f_2^2(x)dx+\cdots+\int_a^b f_N^2(x)dx. \tag{12}$$

这个定理的证明与上面所述(§ 1)n 维空间中同一定理的证明没有什么不同。

直到现在我们还没有给出确切的定义，什么函数应当看成希尔伯特空间中的向量。作为这种函数，当取所有使得积分 $\int_a^b f^2(x)dx$ 有意义的函数。可以设想限制在连续函数上是很自然的，对于连续函数 $\int_a^b f^2(x)dx$ 显然永远存在。假如在广义的意义上理解积分 $\int_a^b f^2(x)dx$，例如在所谓勒贝格积分(参见第十五章)意义上，则希尔伯特空间理论才获得最大的完整性和自然性。

积分(以及相应地考虑的函数类)概念的这个推广对于泛函分析是如此必要，正好像严格的实数理论之对于微积分基础. 就这样，在二十世纪初创造的、与实变函数论发展相关的、平常积分概念的推广，对于泛函分析及与它有关的数学分支显得非常重要.

§3. 依直交函数系的分解

直交函数系的定义和例 假如在平面上取任何二互相直交具单位长的向量 $\boldsymbol{e}_1$ 及 $\boldsymbol{e}_2$(图7)，则在同一平面上任意向量 $\boldsymbol{f}$ 可以依照这两向量的方向分解，就是把它表作下列形式:

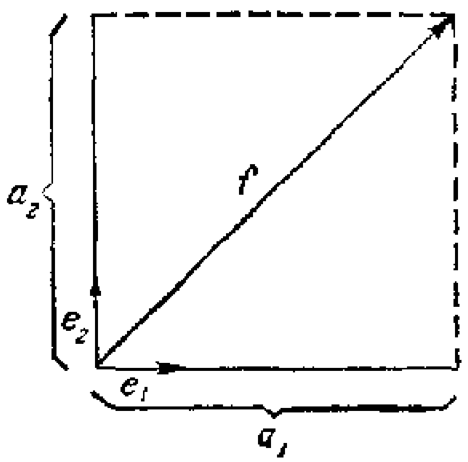


图 7

$$\boldsymbol{f}=a_1\boldsymbol{e}_1+a_2\boldsymbol{e}_2,$$

这里 a_1 及 a_2 是数，等于向量在轴 $\boldsymbol{e}_1$ 及 $\boldsymbol{e}_2$ 的方向上的投影. 因为 $\boldsymbol{f}$ 在轴上的投影等于 $\boldsymbol{f}$ 的长和 $\boldsymbol{f}$ 与轴间角度的余弦的乘积，所以回忆一下纯积定义，我们可写

$$a_1=(\boldsymbol{f},\ \boldsymbol{e}_1),\quad a_2=(\boldsymbol{f},\ \boldsymbol{e}_2).$$

类似地，假如在三维空间中取任何三个互相直交具单位长的向量 $\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \boldsymbol{e}_3$，则在此空间中任意向量 $\boldsymbol{f}$ 可表作下列形式:

$$\boldsymbol{f}=a_1\boldsymbol{e}_1+a_2\boldsymbol{e}_2+a_3\boldsymbol{e}_3,$$

此处 $$a_k=(\boldsymbol{f},\ \boldsymbol{e}_k)\quad(k=1,\ 2,\ 3).$$

在希尔伯特空间中也可考虑这空间中的两两直交的向量系，即函数 $\varphi_1(x),\ \varphi_2(x),\ \cdots,\ \varphi_n(x),\ \cdots$.

函数的这种系称为直交函数系，直交系在分析学中起很大作用. 我们在数学物理、积分方程、逼近计算、实变函数论等等不同的问题中遇到它们. 涉及这种系的概念的整理及统一是一种刺激，这种刺激在二十世纪初导致希尔伯特空间一般概念的创立.

我们来给一个确切的定义. 函数系

$$\varphi_1(x),\ \varphi_2(x),\ \cdots,\ \varphi_n(x),\ \cdots$$

称为直交的，假如这个系的任二函数相互直交，即假如

$$\int_a^b \varphi_i(x)\varphi_k(x)dx=0 \quad 对于\ i\neq k. \tag{13}$$

在三维空间中我们要求系中向量的长度等于1. 回忆一下向量的长的定义，我们看出，在希尔伯特空间中这个条件可写成：

$$\int_a^b \varphi_k^2(x)dx=1. \tag{14}$$

满足条件(13)及(14)的函数系称为直交且就范的。

下面举出这种函数系的若干例子.

1. 在区间$(-\pi,\ \pi)$上考虑函数序列

$$1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \cdots,\ \cos nx,\ \sin nx,\ \cdots.$$

这序列中每两函数相互直交. 这可由相应的积分的简单计算来核验. 在希尔伯特空间中，向量的长的平方是函数平方的积分. 依此，序列

$$1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \cdots,\ \cos nx,\ \sin nx,\ \cdots$$

中的向量的长的平方是积分

$$\int_{-\pi}^{\pi} dx=2\pi,\ \int_{-\pi}^{\pi}\cos^2 nx\,dx=\pi,\ \int_{-\pi}^{\pi}\sin^2 nx\,dx=\pi,$$

就是说，我们的向量序列是直交的，但不是就范的. 序列中的第一个向量的长等于$\sqrt{2\pi}$，而所有其余的具有长度$\sqrt{\pi}$. 将每个向量除以它的长，我们就得到直交就范三角函数系

$$\frac{1}{\sqrt{2\pi}},\ \frac{\cos x}{\sqrt{\pi}},\ \frac{\sin x}{\sqrt{\pi}},\ \frac{\cos 2x}{\sqrt{\pi}},\ \frac{\sin 2x}{\sqrt{\pi}},\ \cdots,$$

$$\frac{\cos nx}{\sqrt{\pi}},\ \frac{\sin nx}{\sqrt{\pi}},\ \cdots$$

这个系在历史上是最早的和最重要的直交系的例子之一. 它出现在欧拉、伯努利、达朗贝尔的工作中，与弦的振动问题有联系. 它的研究在整个分析学[1]的发展中起根本的作用.

---

1) 参见第十二章(第二卷)§1.

直交三角函数系与振动弦的问题相关联地出现不是偶然的．关于介质的微小振动的每个问题化归为描述所给的系的所谓固有振动的某直交函数系(参见§4)．例如与关于球的振动问题相联系出现所谓球函数，与关于圆膜或圆柱的振动问题相联系出现所谓圆柱函数等等．

2．可以举直交函数系的一例，它的每个函数表现为多项式．作为这种例出现的有勒让德多项式序列

$$P_n(x)=\frac{1}{2^n n!}\frac{d^n(x^2-1)^n}{dx^n},$$

就是说，$P_n(x)$是(除去一常数因子不计)$(x^2-1)^n$的n阶导数．写出这序列的开头若干个多项式：

$$P_0(x)=1;$$

$$P_1(x)=x;$$

$$P_2(x)=\frac{1}{2}(3x^2-1);$$

$$P_3(x)=\frac{1}{2}(5x^3-3x).$$

显然，普遍地$P_n(x)$是n次多项式．让读者自己证明，这些多项式是区间$(-1, 1)$上的直交序列．

著名的俄国数学家П. Л. 车比雪夫在十九世纪后半叶发展了直交多项式(也就是所谓具有权的直交多项式)的一般理论．

依函数直交系的分解　和在三维空间中每个向量可以表成相互直交具单位长的三个向量的线性组合形式

$$\boldsymbol{f}=a_1\boldsymbol{e}_1+a_2\boldsymbol{e}_2+a_3\boldsymbol{e}_3$$

相类似，在泛函空间中发生将任意函数依直交就范函数系展为级数的问题，就是把函数f表成下列形式：

$$f(x)=a_1\varphi_1(x)+a_2\varphi_2(x)+\cdots+a_n\varphi_n(x)+\cdots \tag{15}$$

的问题．在这里，级数(15)到函数f的收敛性是在希尔伯特空间中元素间距离的意义上来理解的．这意思就是从级数部分和$S_n(t)=\sum_{k=1}^{n}a_k\varphi_k(t)$到函数$f(t)$的均方差趋于0(当$n\to\infty$)，即

$$\lim_{n\to\infty}\int_a^b [f(t)-S_n(t)]^2 dt=0. \qquad (16)$$

这样的收敛性称为"平均收敛性"。

依这个或那个函数直交系的分解常常在分析学中遇得到，这样的分解常常是解决数学物理问题的重要方法。例如，若直交系是区间$(-\pi, \pi)$上的三角函数系

$$1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \cdots,\ \cos nx,\ \sin nx,\ \cdots,$$

则这样的分解就是函数的成为三角级数的古典分解[1]：

$$f(x)=a_0+a_1\cos x+b_1\sin x+a_2\cos 2x+b_2\sin 2x+\cdots.$$

假定分解(15)对于希尔伯特空间中的任意函数f为可能的，我们来找出这一分解的系数a_n。为此，将等式的两端与我们的系中同一函数φ_m求纯积。我们得到等式

$$(f,\ \varphi_m)=a_1(\varphi_1,\ \varphi_m)+a_2(\varphi_2,\ \varphi_m)+\cdots+a_m(\varphi_m,\ \varphi_m)$$
$$+a_{m+1}(\varphi_{m+1},\ \varphi_m)+\cdots,$$

由于当$m\neq n$时有$(\varphi_m,\ \varphi_n)=0$，又$(\varphi_m,\ \varphi_m)=1$，由上式可知，系数$a_m$的值为

$$a_m=(f,\ \varphi_m)\quad (m=1,\ 2,\ \cdots).$$

我们看出，像在普通三维空间中一样(参见这段的开始)，系数a_m等于向量f在向量φ_m方向上的投影。

借助于纯积的定义得到函数$f(x)$依直交就范函数系$\varphi_1(x)$，$\varphi_2(x)$，$\cdots$，$\varphi_n(x)$，$\cdots$的分解：

$$f(x)=a_1\varphi_1(x)+a_2\varphi_2(x)+\cdots+a_n\varphi_n(x)+\cdots, \qquad (17)$$

它的系数由下列公式定义：

$$a_m=\int_a^b f(t)\varphi_m(t)dt. \qquad (18)$$

作为例，考虑上面所引直交就范三角函数系：

$$\frac{1}{\sqrt{2\pi}},\ \frac{\cos x}{\sqrt{\pi}},\ \frac{\sin x}{\sqrt{\pi}},\ \frac{\cos 2x}{\sqrt{\pi}},\ \frac{\sin 2x}{\sqrt{\pi}},\ \cdots,$$

---

1) 将振动分解成为调和的组成部分，这样的分解常在许多物理问题中遇到。参见第六章(第二卷)§5。

则
$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos nx+b_n\sin nx,$$
此处
$$a_0=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\,dx,\quad a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos nx\,dx,$$
$$b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin nx\,dx.$$
我们对于函数的三角级数展开的系数计算已有公式，只要这个展开是可能的[1].

我们决定了函数 $f(x)$ 依直交函数系的展开的系数形式——在这种展开是存在的假定下．但是无穷直交函数系 $\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots$ 对于希尔伯特空间中任意函数依这个系展开可以显现为不充分．为使这种展开是可能的，直交函数系应满足补充的条件，所谓完备性条件.

直交函数系称为**完备的**，假如对于它不可能再加一个不恒等于零而与系中的所有函数直交的函数.

容易举出不完备的直交系例子．为此，取任何直交系，例如还取三角函数系而去掉这个系中一个函数，例如 $\cos x$．所余的无穷函数系

$$1,\ \sin x,\ \cos 2x,\ \sin 2x,\ \cdots,\ \cos nx,\ \sin nx,\ \cdots$$

像先前一样也是直交的，但不是完备的，因为被我们去掉的函数 $\cos x$ 与系中所有的函数直交.

假如函数系不完备，则不可能将希尔伯特空间中的一切函数依它展开．事实上，假如我们企图把一非零而与系中的所有函数直交的函数 $f_0(x)$ 依这样的系展开，则由于公式(18)，所有的系数都变成等于 0，而同时函数 $f_0(x)$ 又不等于 0.

有以下的定理：假如在希尔伯特空间中给了完备直交就范函数系 $\varphi_1(x), \varphi_2(x), \cdots, \varphi_n(x), \cdots$，则所有的函数 $f(x)$ 都可依这个系中的函数展开成级数[2]

$$f(x)=a_1\varphi_1(x)+a_2\varphi_2(x)+\cdots+a_n\varphi_n(x)+\cdots,$$

1) 关于三角级数参见第十二章(第二卷)§7.

2) 和前面一样，这个级数与它的和的关系按公式(16)的意义来了解.

在这里展开式系数 a_n 等于向量在直交就范系中的元素上投影：

$$a_n=(f,\ \varphi_n)=\int_a^b f(x)\varphi_n(x)dx.$$

§2中的毕达哥拉斯定理(在希尔伯特空间中)使得我们能找到系数 a_k 与函数 $f(x)$ 间的有趣的关系．用 $r_n(x)$ 表 $f(x)$ 与它的级数的首 n 项和的差，即

$$r_n(x)=f(x)-[a_1\varphi_1(x)+\cdots+a_n\varphi_n(x)].$$

函数 $r_n(x)$ 与 $\varphi_1(x)$，$\varphi_2(x)$，…，$\varphi_n(x)$ 直交．事实上，核验出来它与 $\varphi_1(x)$ 直交，即 $\int_a^b r_n(x)\varphi_1(x)dx=0$．我们有

$$\int_a^b r_n(x)\varphi_1(x)dx=\int_a^b[f(x)-a_1\varphi_1(x)-a_2\varphi_2(x)-\cdots-a_n\varphi_n$$

$$-a_n\varphi_n(x)]\varphi_1(x)dx=\int_a^b f(x)\varphi_1(x)dx-a_1\int_a^b\varphi_1^2(x)dx\ ^{1)}.$$

因为 $a_1=\int_a^b f(x)\varphi_1(x)dx$ 而 $\int_a^b\varphi_1^2(x)dx=1$，故推知

$$\int_a^b r_n(x)\varphi_1(x)dx=0.$$

如此，在等式中

$$f(x)=a_1\varphi_1(x)+a_2\varphi_2(x)+\cdots+a_n\varphi_n(x)+r_n(x)\tag{19}$$

右端各个项相互直交．意谓，由于在§2陈述的毕达哥拉斯定理，向量 $f(x)$ 的长度平方等于等式(19)右端各个项的长度平方和，即

$$\int_a^b f^2(x)dx=\int_a^b[a_1\varphi_1(x)]^2dx+\cdots+\int_a^b[a_n\varphi_n(x)]^2dx+\int_a^b r_n^2(x)dx.$$

因为函数系 φ_1，φ_2，…，φ_n，…是就范的[等式(14)]，故

$$\int_a^b f^2(x)dx=a_1^2+a_2^2+\cdots+a_n^2+\int_a^b r_n^2(x)dx.\tag{20}$$

级数 $\sum_{k=1}^{\infty}a_k\varphi_k(x)$ 平均收敛．这就是说

$$\int_a^b[f(x)-a_1\varphi_1(x)-\cdots-a_n\varphi_n(x)]^2dx\to 0,$$

即

$$\int_a^b r_n^2(x)dx\to 0.$$

但从公式(20)我们得到等式

$$\sum_{k=1}^{\infty}a_k^2=\int_a^b f^2(x)dx\ ^{2)},\tag{21}$$

---

1) 其他的积分等于零，因为函数 $\varphi_n(x)$ 相互直交．

2) 几何上这意谓希尔伯特空间向量的长度平方等于在相互直交方向完备系上向量的投影平方和．

它断定了函数平方的积分等于函数依封闭(完备——译者)直交函数系的展开的系数平方和. 假如条件(21)对于希尔伯特空间中的任意函数都成立，则它称为完备性条件.

再来注意下面的重要情况. 怎样的数 a_k 可以成为希尔伯特空间中的函数的分解的系数?等式(21)断言,为了这,级数 $\sum_{k=1}^{\infty} a_k^2$ 必须收敛. 原来这个条件是充分的,即，为了使数 a_k 序列是依希尔伯特空间中直交函数系的分解的系数序列,必要且充分条件是:级数 $\sum_{k=1}^{\infty} a_k^2$ 收敛.

注意，假如定义希尔伯特空间为一切依勒贝格意义平方可积的函数的总体(参见 227 页)，则上述这个根本定理成立. 假如在希尔伯特空间中例如仅以连续函数为限，则什么数可作为分解的系数，关于这问题的解答是不必要地紧复的.

这里所引进的考虑，是导致在希尔伯特空间的定义中使用推广意义上的(依勒贝格)积分的必要理由之一.

§4. 积 分 方 程

在这一节中我们介绍给读者泛函分析的重要分支之一，也是历史上最早的分支之一，就是在泛函分析进一步发展中起如此根本作用的积分方程的理论. 在积分方程理论的发展中，除去数学的内部要求[例如偏微分方程边界问题(第二卷第六章)]外，还有大量的各种物理问题. 除微分方程而外，在二十世纪积分方程显现为各种物理问题的数学研究的重要工具之一. 在这节中我们陈述积分方程论中某些知识. 我们在这里陈述的那些事实密切地联系着并在很大程度上(直接或间接)产生于弹性体系微小振动的研究.

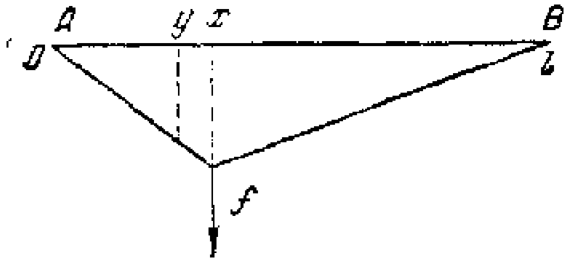


图 8

关于弹性体系微小振动的问题 回到在§2中考虑的关于微小振动的问题. 试求描述这种振动的方程. 为了假设的简单，我们以线形弹性体系的振动为对象，譬如说，长 l 的弦(图 8)或弹性枢轴(图 9)可作为这种体

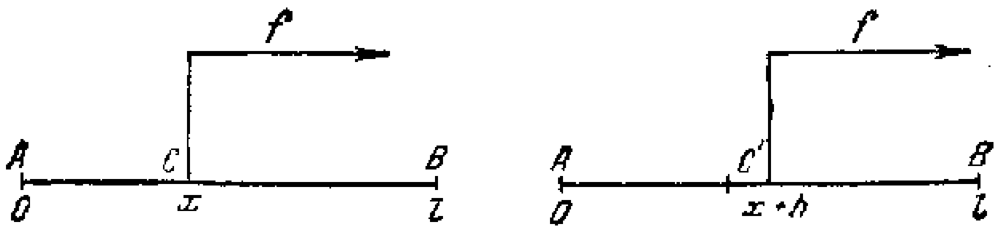


图 9

系的例. 假设在平衡位置时我们的弹性体位置在 Ox 轴的线段 Ol 上. 在点 x 处施加单位力. 在这一力的作用下, 体系的所有点得到某一离差. 在点 y 处产生的离差(图 8)用 $k(x, y)$ 来表示.

函数 $k(x, y)$ 显现为两点函数: 点 x, 在该处施力, 和点 y, 在该处我们计量离差. 它被称为影响函数.

从能量守恒定律, 可以导引出影响函数 $k(x, y)$ 的重要性质, 即所谓互易原理: 在点 x 处施力在点 y 处产生的离差等于在点 y 处施同一大小的力在点 x 处产生的离差. 换言之, 这意谓

$$k(x, y)=k(y, x). \tag{22}$$

例如, 对于弹性枢轴沿轴的方向的振动试求影响函数[在图 8 上画出了其他的(横断方向的)位移]. 考虑长为 l 的枢轴 AB, 固定其端点(图 9). 在点 C 处施以沿方向 B 作用的力 f. 在这一力作用下枢轴变形了, 而点 C 移至位置 C', 点 C 的位移大小用 h 表示. 先求 h, 借助于 h 我们又可以求在任意点 y 处的位移. 为此要利用虎克定律, 定律断定, 力与相对伸长(即位移与长的比)成比例. 类似的关系对于压缩也成立.

在力 f 作用下, 枢轴的 AC 部分伸长了. 由于这力产生的反作用表以 T_1. 在这时候部分 CB 被压缩了, 产生了反作用 T_2. 根据虎克定律

$$T_1=\varkappa\frac{h}{x},\ T_2=\varkappa\frac{h}{l-x},$$

此处 $\varkappa$ 是表示枢轴的弹性特征的比例系数. 根据作用在点 C 处诸力平衡的条件, 我们有

$$f=\varkappa\frac{h}{x}+\varkappa\frac{h}{l-x},\ 即\ f=\frac{\varkappa lh}{x(l-x)}.$$

由此有 $$h=\frac{f}{\varkappa l}x(l-x).$$

为了求在位于线段 AC 上某一点 y $(y<x)$ 处产生的位移，我们指出，从虎克定律推知，在枢轴伸长时相对的伸长(即点的位移与由点到不动的端点的比率)与点的位置无关．用 k 表点 y 的位移，令在点 x 及 y 处的相对位移相等，则得

$$\frac{k}{y}=\frac{h}{x},$$

由此当 $y<x$ 时

$$k=h\frac{y}{x}=\frac{f}{\varkappa l}y(l-x).$$

类似地，假如点 y 位于线段 CB 上 $(y>x)$，得

$$k=h\frac{l-y}{l-x}=\frac{f}{\varkappa l}x(l-y).$$

忆及影响函数 $k(x, y)$ 是在点 x 处施加单位力的作用下点 y 处的离差，得知当弹性枢轴沿轴的方向振动时影响函数具有下列形式：

$$k(x, y)=\begin{cases}\dfrac{1}{\varkappa l}y(l-x) & 于\ y<x\ 时,\\[2ex] \dfrac{1}{\varkappa l}x(l-y) & 于\ y>x\ 时.\end{cases}$$

可用或多或少相似的方式求得关于弦的影响函数．假如弦的张力为 T 而长为 l，则在点 x 处施加单位力的作用下，弦获得在图8上描绘的形式，且在点 y 处的位移 $k(x, y)$ 由下列公式给出：

$$k(x, y)=\begin{cases}\dfrac{1}{Tl}x(l-y) & 于\ x<y\ 时,\\[2ex] \dfrac{1}{Tl}y(l-x) & 于\ x>y\ 时,\end{cases}$$

这与前面得到的关于枢轴的影响函数相一致．

在以下情形中，通过影响函数可以表现体系对平衡位置的离差，即在体系上作用着连续分布的以 $f(y)$ 为密度的力的情形．因为在长为 Δy 的区间上作用着力 $f(y)\Delta y$，我们可近似地把它看成集中在点 y 处，所以在这力作用下在点 x 处产生离差

$k(x, y)f(y)\Delta y$. 在整个荷载作用下离差可近似地看作等于和数

$$\sum k(x, y)f(y)\Delta y.$$

在 $\Delta y \to 0$ 的极限情形下，可以看出，在沿着体系分布的力 $f(y)$ 作用下，点 x 处的离差 $u(x)$ 由下列公式给出：

$$u(x)=\int_a^b k(x, y)f(y)dy. \tag{23}$$

假设我们的弹性体系没有受到外力作用．假如把它从平衡位置移动开来，则它开始运动．这种运动称为体系的自由振动．

现在借助于影响函数 $k(x, y)$ 来描述所考虑的弹性体系所服从的方程．为此，用 $u(x, t)$ 来表示点 x 在时刻 t 距平衡位置的离差．于是，点 x 处在时刻 t 的加速度等于 $\dfrac{\partial^2 u(x, t)}{\partial t^2}$.

假如 ρ 是物体的线密度，即 ρdy 是长度元素 dy 的质量，则依照力学基本规律我们获得运动方程，只要在公式(23)中用质量与加速度的乘积 $\dfrac{\partial^2 u(x, t)}{\partial t^2}\rho dy$ 加以变号来代替力 $f(y)dy$.

照这样，自由振动的方程具有形式

$$u(x, t)=-\int_a^b k(x, y)\frac{\partial^2 u(y, t)}{\partial t^2}\rho dy.$$

在振动理论中，所谓弹性体系调和振动起很大作用，这振动即是这样的运动，对于它

$$u(x, t)=u(x)\sin \omega t,$$

它们标示下述特性：每个固定点作出调和振动(依正弦规律运动)，以 ω 为频率，这个频率对于所有点 x 是相同的．

我们还看到，每个自由振动是调和振动的组成部分．

命
$$u(y, t)=u(y)\sin \omega t,$$
将它代入自由振动的方程中并消去 $\sin \omega t$．于是我们得到决定函数 $u(x)$ 的下列方程：

$$u(x)=\rho\omega^2\int_a^b k(x, y)u(y)dy. \tag{24}$$

这样的方程称为关于函数 $u(x)$ 的齐次积分方程．

显然，对任意的 ω，方程(24)有我们不感兴趣的解 $u(x)=0$，它与静止状态相应. 使方程(24)有异于零解的那些 ω 值称为体系的固有值.

因为不是对于每个 ω 值都有异于零的解，体系能作自由振动仅依赖于特定的频率，其中最小的称为基本音频，而其余的称为泛音频.

可以证明，每个体系有固有频率的无穷序列，就是所谓频率谱

$$\omega_1,\ \omega_2,\ \cdots,\ \omega_n,\ \cdots.$$

方程(24)的非零解 $u_n(x)$，对应于频率 ω_n，给出相应的固有振动形式.

例如，假设弹性体系表现为弦，紧拉在点 O 及 l 之间且固定在这两点，则对于这样的体系可能的固有振动等于

$$a\frac{\pi}{l},\ 2a\frac{\pi}{l},\ 3a\frac{\pi}{l},\ \cdots,\ na\frac{\pi}{l},\ \cdots,$$

此处 a 是系数，与密度及弦的张力有关，即 $a=\sqrt{\frac{T}{\rho}}$. 基本音谱为 $\omega_1=a\frac{\pi}{l}$，而泛音频为 $\omega_2=2\omega_1,\ \omega_3=3\omega_1,\ \cdots,\ \omega_n=n\omega_1$. 相应的调和振动形式由下列等式给出：

$$u_n(x)=\sin\frac{na\pi}{l}x.$$

而对于 $n=1,\ 2,\ 3,\ 4$ 在图 10 中有描绘的图形.

我们已考虑过弹性体系的自由振动. 假如在运动的同时有调和的外力作用在弹性体系上，则当我们决定在这一力作用下的调和振动时，对于函数 $u(x)$ 得到所谓非齐次积分方程

$$u(x)=\rho\omega^2\int_a^b k(x,\ y)u(y)dy+h(x). \tag{25}$$

积分方程性质 在上面我们已看到了积分方程的例：

$$f(x)=\lambda\int_a^b k(x,\ y)f(y)dy \tag{26}$$

及

$$f(x)=\lambda\int_a^b k(x,\ y)f(y)dy+h(x), \tag{27}$$

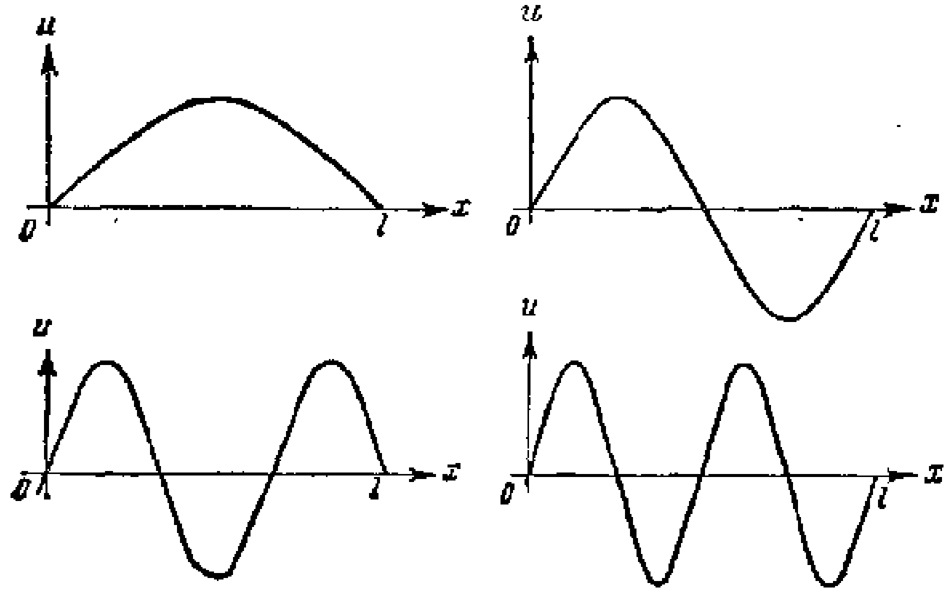


图 10

其中第一个是从关于弹性体系自由振动问题的解决得来的，而第二个是从强迫振动，即在外力作用下振动的研究中得来的.

在这些方程中未知函数为函数 $f(x)$. 已知的函数 $k(x, y)$ 称为积分方程的核.

方程(27)称为非齐次线性积分方程，而方程(26)称为齐次的. 它是于 $h(x)=0$ 时从非齐次方程得到的.

显然，齐次方程永远有零解，即解 $f(x)=0$. 在非齐次与齐次积分方程解之间存在着密切联系. 指出下列的定理作为例：假如齐次积分方程仅有零解，则对应的非齐次方程对于所有的函数 $h(x)$ 是可解的.

假如对于 λ 的某个值齐次方程有不恒等于 0 的解 $f(x)$，则这个 λ 值称为固有值，而对应的解 $f(x)$ 称为固有函数. 我们从上面看到，当积分方程描述弹性系统的自由振动时，则固有值与体系的振动频率密切相关(即 $\lambda=\rho\omega^2$). 固有函数给出对应的调和振动的形式.

对于振动问题从能量守恒定律推出

$$k(x, y)=k(y, x). \tag{28}$$

满足条件(28)的核称为对称的.

对于具对称核的方程，固有函数及固有值拥有一系列的重要

性质．可证明：这样的方程永远有一系列的实固有值

$$\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n,\ \cdots.$$

对每个固有值有一个或若干个固有函数相应．同时对应于不同的固有值的固有函数永远是相互直交的[1]．

照这样，对于每个具对称核的积分方程固有函数系是直交函数系．一个问题发生：当这个系完备，也就是当希尔伯特空间中的每个函数可依积分方程固有函数系展开成级数，特别是假如方程

$$\int_a^b k(x,\ y)f(y)dy=0 \tag{29}$$

仅当 $f(y)\equiv 0$ 时满足，则积分方程

$$\lambda\int_a^b k(x,\ y)f(y)dy=f(x)$$

的固有函数系是直交系[2]．

照这样，每个平方可积函数 $f(x)$ 可在这情形中依固有函数展开成级数．从考虑这些或那些积分方程中，为了各种重要直交系的封闭性，我们获得普遍而有力的方法，即函数依直交函数展开成级数的可能性的证明．用这些方法可证明三角函数系、圆柱函数系、球函数系及其他重要函数系的完备性．

在振动情形中，任意函数可依固有函数展开成级数，这一现象意味着任意振动可展开为调和振动的和．这样的展开成为一种方法，广泛地应用于关于力学及物理学不同领域中的振动（弹性体振动、声学的振动、电磁波等等）问题的解决中．

线性积分方程理论的发展，对于线性运算子一般理论的创立起了推动的作用，线性积分方程理论有机地进入到后一理论中．在近数十年中，线性运算子理论一般方法也有力地促进了积分方程

1) 最后的论断将在下一节（第 246 页）中证明．

2) 在 $k(x,\ y)$ 是弹性体系的影响函数时，条件(29)有着普通的物理意义．在这件事情上［参考公式(23)］，我们已见到，在沿体系分布的力 $f(y)$ 作用下，体系距平衡位置的离差由公式

$$u(x)=\int_a^b k(x,\ y)f(y)dy$$

表达．照这样，条件(29)意味着：每个异于零的力把体系自平衡位置带出来．

理论进一步的发展.

§5. 线性运算子及泛函分析进一步的发展

在前一节中我们看到，弹性体系振动问题导致寻找积分方程固有值及固有函数的问题. 注意这些问题也可以与线性微分方程[1]固有值及固有函数的寻求相联系，许多其它物理问题也归结为线性微分或积分方程固有值及固有函数的计算问题.

再举一个例. 在现代的无线电技术中广泛地应用一种传播高频电磁振荡的所谓波导，就是空心的金属管，在它里面传播电磁波. 大家都知道，沿着波导只能传播波长不太长的电磁振荡. 临界波长的寻求化归为某个微分方程固有值的问题.

除此之外，在线性代数、常微分方程理论、稳定性问题等中也会遇到固有值问题.

从统一的一般观点考虑所有这些相互类似的问题的必要性产生了. 这样的观点显现为线性运算子的一般理论. 与在不同的具体问题中固有函数及固有值有关的许多问题，只有在运算子一般理论的观点下才能彻底理解. 照这样，在这个方向和一系列其它方向中，运算子一般理论本身成为在产生它的领域的研究中很有成效的工具.

在运算子理论的进一步发展过程中，在量子力学里广泛利用运算子理论中的方法表现为本质的阶段. 所谓自共轭运算子理论已成为量子力学的基本数学工具. 在量子力学中产生的数学问题的提法已经成为且将继续成为对于泛函分析进一步发展的刺激.

在微分和积分方程中，已经证实，对于这些方程解的实用逼近法的发展，运算子观点是非常有用的.

运算子理论的基本观点 转过来阐述与运算子理论有关的基本定义和事实.

---

1) 参见第二卷第六章 §5.

在分析学中我们遇到了函数概念．在最简单形式中，它是对应关系：对于每个数 x(自变数的值)有一数 y(函数值)与它相对应．随着分析学进一步的发展，产生了考虑属于更普遍类型的对应关系的需要．

例如，在变分学(第二卷第八章)中考虑到更一般的对应关系：对于每个函数有数来与它对应．假如对每个函数有数来与它相应，那么我们说，我们得到一个泛函数．对每个函数 $y=f(x)$ $(a\leqslant x\leqslant b)$，令被它描绘的曲线轨迹的长度来与它相对应，可做为泛函数的例子．假如对每个函数 $y=f(x)$ $(a\leqslant x\leqslant b)$，令它的定积分 $\int_a^b f(x)\,dx$ 与它相对应，我们就得到泛函数的又一个例子．

假如把 $f(x)$ 看成无穷维空间中的点，那么泛函数是无穷维空间中点的函数而不是别的．从这个观点看，变分学中所研究的就是关于无穷维空间点的函数的极大或极小值的寻求问题．

为了定义什么是连续泛函数，必须定义无穷维空间中两点邻近的意义是什么．在§2中我们给出了两个函数[无穷维空间中的点 $f(x)$ 及 $g(x)$]间距离为 $\sqrt{\int_a^b [f(x)-g(x)]^2\,dx}$．无穷维空间中距离的这种定义方式常被使用，但当然这并不是唯一可能的定义方式．在其它问题中表明函数间距离的其它定义方式更好．举一个例：在函数逼近理论(参见第二卷第十二章§3)中的问题里，函数间距离的定义由标示两个函数 $f(x)$ 及 $g(x)$ 相近度的下列公式给出：

$$\max|f(x)-g(x)|.$$

在变分学中研究泛函数时，要用到函数间距离的其他定义方式．

函数间距离不同的定义方式把我们引到不同的无穷维空间．

这样，不同的无穷维(泛函)空间由函数储备及其间距离的定义互相区别．例如，假定取所有的平方可积函数的总体并定义距离为 $\sqrt{\int_a^b [f(x)-g(x)]^2\,dx}$，那么我们就得到在§2中引入的希

尔伯特空间；假定我们取所有的连续函数总体并定义距离为 $\max|f(x)-g(x)|$，那么我们得到所谓空间(C)。

当考虑积分方程时，我们碰到形如

$$g(x)=\int_a^b k(x, y)f(y)dy$$

的表式．对于给定的核 $k(x, y)$，所写的等式显示一个规律：对于每个函数 $f(x)$ 有另一函数 $g(x)$ 与之相对应．

这样的对应，对于一个函数 f 使另一函数 g 来对应，称为**运算子**．

假如给了一个规律，按照这个规律，对每个函数 f，有函数 g 与之相对应，那么就说，我们得到希尔伯特空间中线性运算子 A．不是对希尔伯特空间中的所有函数的对应关系也可给定．在这情形中，使得函数 $g=Af$ 存在的那些 f 的集合称为运算子 A 的**定义域**（类似于普通分析学中函数的定义域）．这个对应关系和平常一样表记为

$$g=Af. \tag{30}$$

运算子的线性意味着：对于函数 f_1 及 f_2 的和有 Af_1 与 Af_2 的和与之相对应，而对于函数 f 被数 λ 相乘有函数 λAf 与之相对应，即

$$A(f_1+f_2)=Af_1+Af_2 \tag{31}$$

及

$$A(\lambda f)=\lambda Af. \tag{32}$$

有时对线性运算子同时要求连续性，就是，要求在函数序列 f_n 到函数 f 的收敛下序列 Af_n 收敛于 Af。

举若干线性运算子的例．

1° 对每个函数 $f(x)$，使函数 $g(x)=\int_a^x f(t)dt$，即函数 f 的不定积分与之相对应．这个运算子的线性从积分的下列通常性质引出，即从这引出：和的积分等于积分的和，又常数因子可以移到积分符号外面．

2° 对于每个可微分函数 $f(x)$，使它的导数 $f'(x)$ 与之相对

应．这个运算子通常用字母 D 来表达，即

$$f'(x)=Df(x).$$

注意这个运算子不是对于希尔伯特空间中的所有函数有定义，而仅对于有属于希尔伯特空间的导数的函数有定义．这些函数组成(如上述)所给运算子的定义域．

3° 例 1° 及 2° 是在希尔伯特空间中线性运算子的例子．我们在本书其它章节里曾遇到过有穷维空间中线性运算子的例子．在第三章(第一卷)中研究过仿射变换．假如平面上或空间中仿射变换使坐标原点留在原处，那么它是在二维的或相应地在三维的空间中线性运算子的例子．在第十六章中引进了 n 维空间中的线性变换，它是 n 维空间中的线性运算子．

4° 在积分方程中，我们已在实质上遇到非常重要且在分析学中有广泛应用的泛函空间中的线性运算子——所谓积分运算子类．给定一个确定的函数 $k(x, y)$，则公式

$$g(x)=\int_a^b k(x, y)f(y)dy$$

对于每个函数 f 使一函数 g 与之相对应．符号上我们可以把这个变换照以下形式写出：

$$g=Af.$$

在这个情形中，运算子 A 称为积分运算子．还可以给出许多积分运算子的重要例子．

在 §4 中我们说到非齐次积分方程

$$f(x)=\lambda\int_a^b k(x, y)f(y)dy+h(x).$$

用运算子理论的记法，这个方程可改写为

$$f=\lambda Af+h, \tag{33}$$

此处 λ 是已给的数，h 是已给的函数(无穷维空间中的向量)，f 是待求的函数．仍用这些记法，齐次方程可改写如下：

$$f=\lambda Af. \tag{34}$$

关于积分方程的古典理论，例如，在 §4 中陈述过的关于非齐次及

相应的齐次积分方程可解性间关系的定理，并不是对于所有运算子方程正确的．然而可以指出来若干一般条件，加于运算子 A 之上．在这些条件之下，定理成为真确的．

这些条件用拓扑学的词句陈述出来而总结为：运算子 A 把单位球体(就是长度不超过 1 的向量总体)变换到紧集中．

运算子的固有值及固有向量 由关于振动的问题把我们引到关于积分方程固有值及固有函数的问题，这些问题我们已照以下方式陈述过：求 λ 的值使得有异于零的函数 f，满足方程

$$f(x)=\lambda\int_a^b k(x, y)f(y)dy.$$

像先前一样，这个方程可改写成

$$f=\lambda Af$$

或

$$Af=\frac{1}{\lambda}f. \tag{35}$$

今后将 A 理解为任意线性运算子．于是满足等式(35)的向量 f 称为运算子 A 的固有向量而数 $\frac{1}{\lambda}$ 为相应的固有值．

因为向量 $\frac{1}{\lambda}f$ 在方向上与向量 f 一致（与 f 仅相差一数值因子)，所以，寻求固有向量的问题还可以陈述为寻求在变换 A 之下不变它自己的方向的非零向量 f 的问题．

在关于固有值问题上的这样的观点，使得我们能够把关于积分方程(假如 A 是积分运算子)、微分方程(假如 A 是微分运算子)的固有值问题和关于线性代数中[假如 A 是有穷维空间中线性变换；参见第六章(第二卷)及第十六章]的固有值问题统一起来，对于三维空间情形这个问题已在求椭球体的所谓主轴中遇到．

在积分方程的情形，固有函数及固有值的一系列重要性质(例如固有值的实数性，固有函数的直交性等)表现为核的对称性——就是等式 $k(x, y)=k(y, x)$ 的推论．

对于希尔伯特空间中任意的线性运算子 A，运算子的所谓自共轭性表现为这个性质的类似．

在一般情形中，运算子 A 的自共轭性条件如下：对任意两元素 f_1 及 f_2，等式

$$(Af_1,\ f_2)=(f_1,\ Af_2)$$

成立，此处 $(Af_1,\ f_2)$ 表向量 Af_1 与向量 f_2 的纯积.

运算子的自共轭性在力学问题中通常表现为能量守恒定律的推论. 因之，这个条件对于(譬如说)与振动[此时没有能量耗损(散逸)]有关联的运算子是成立的.

在量子力学中遇到的大多数运算子都是自共轭的.

我们来验证具对称核 $k(x,\ y)$ 的积分运算子是自共轭的. 事实上，在这情形中 Af_1 是函数 $\int_a^b k(x,\ y)f_1(y)dy$. 因之，纯积 $(Af_1,\ f_2)$ 等于这个函数与 f_2 的积的积分，由下列公式给出：

$$(Af_1,\ f_2)=\int_a^b\int_a^b k(x,\ y)f_1(y)f_2(x)dy\,dx.$$

类似地

$$(f_1,\ Af_2)=\int_a^b\int_a^b k(x,\ y)f_2(y)f_1(x)dy\,dx.$$

等式 $(Af_1,\ f_2)=(f_1,\ Af_2)$ 是核 $k(x,\ y)$ 的对称性的直接推论.

任意自共轭运算子拥有一系列重要性质，它们在应用这种运算子来解各种问题当中是有用的. 可以证明：线性自共轭运算子的固有值永远是实的，而且与不同的固有值对应的固有函数相互直交.

举例来证明最后的论断. 设 λ_1 及 λ_2 是运算子 A 的两个不同的固有值，而 f_1 及 f_2 是与它们对应的固有向量. 这意味着

$$\begin{aligned}Af_1&=\lambda_1 f_1,\\ Af_2&=\lambda_2 f_2.\end{aligned}\tag{36}$$

取等式(36)的第一个与 f_2 的纯积，并取第二个与 f_1 的纯积，我们得到

$$\begin{aligned}(Af_1,\ f_2)&=\lambda_1(f_1,\ f_2),\\ (Af_2,\ f_1)&=\lambda_2(f_2,\ f_1).\end{aligned}\tag{37}$$

因为运算子 A 是自共轭的，所以 $(Af_1,\ f_2)=(Af_2,\ f_1)$. 从(37)的第一个等式减去第二个，得

$$0=(\lambda_1-\lambda_2)(f_1,\ f_2).$$

因为 $\lambda_1\neq\lambda_2$，所以 $(f_1,\ f_2)=0$，就是向量 f_1 与 f_2 直交.

自共轭运算子的研究弄明白了许多具体问题和与固有值理论有联系的问题.

我们现在仔细地看看这些问题之中的一个，就是关于在连续谱情形中依固有函数的展开问题.

为了阐明什么是连续谱，再借助于古典的弦振动问题．上面我们已表明，对于长为 l 的弦，振动的固有频率可取一系列的值：

$$a\frac{\pi}{l},\ 2a\frac{\pi}{l},\ \cdots,\ na\frac{\pi}{l},\ \cdots.$$

在数轴 $O\lambda$ 上安置这一系列的点．假如增大弦长 l，则系列中任何两相邻的点的距离就将减小，而它们就将更稠密地填挤在数轴上．在极限情形，当 $l\to\infty$，就是对于无穷长的弦，固有频率就填充半数轴 $\lambda\geqslant 0$．这情形表示体系有连续谱.

我们已经说过，对于长 l 的弦，依固有函数的展开就是依 $n\frac{\pi}{l}x$ 的正弦及余弦的展开，就是展成三角级数：

$$f(x)=\frac{a_0}{2}+\Sigma\, a_n\cos n\frac{\pi}{l}x+b_n\sin n\frac{\pi}{l}x.$$

对于无穷长的弦的情形，可以证明，或多或少任意的函数可依正弦及余弦展开．而既然固有频率现在是连续地分布在数直线上，这展开将不是展开成级数而是展开成所谓傅里叶积分：

$$f(x)=\int_{-\infty}^{\infty}[A(\lambda)\cos\lambda x+B(\lambda)\sin\lambda x]\,d\lambda.$$

展开成傅里叶积分很早就已经知道，且在十九世纪中当解不同的数学物理问题时广泛地被使用过.

而在具连续谱的更一般问题[1)]中，许多与依固有函数展开有关的问题还没有弄清．只有自共轭运算子一般理论的建立，才会使这些问题得到必要的阐明.

还要指出一个以运算子一般理论为基础得到解决的一些古典问题范围．在有能量散逸时振动的考虑引到这些问题.

在这情形中我们仍可找体系的形如 $u(x)\varphi(t)$ 的自由振动．但

1) 非均匀弹性介质振动以及量子力学的许多问题可以作为例子.

是，和没有散逸的振动情形不同，函数 $\varphi(t)$ 不是简单的 $\cos\omega t$ 而是具有形式 $e^{-kt}\cos\omega t$，此处 $k>0$. 照这样，对应的解具有形式 $u(x)e^{-kt}\cos\omega t$. 每个点 x 在这情形中仍作振动（具频率 ω），然而振动表现为衰减的，因为于 $t\to\infty$ 时这些振动的振幅所含的因子 e^{-kt} 趋于 0.

体系的固有振动可以很方便地写作复数的形式：$u(x)e^{-i\lambda t}$，当没有摩擦时此处 λ 是实的，当有摩擦时此处 λ 是复的.

具有能量散逸的体系的振动问题仍然化归成固有振动的问题，但不再是对于自共轭运算子. 对于体系讲，复固有值的存在表现出自由振动衰减的特征.

将运算子理论方法与解析函数论方法的联合使用，M. B. 凯尔迪什在 1950—1951 年研究了这类问题，关于这类问题证明了固有函数系统的完备性.

泛函分析与数学其他分支及量子力学的联系 上面我们已提到，量子力学的创立表现为泛函分析发展过程中具有决定性的一个阶段.

就象微积分的兴起在十八世纪中为力学及古典物理学的需要所统治一样，泛函分析的发展已经并且将继续在现代物理学主要是量子力学最强有力的影响下出现. 本质上与泛函分析有关的数学分支已成为量子力学的基本数学工具. 我们只扼要地指出这里所涉及的联系，量子力学基础的陈述则将超出本书的范围.

在量子力学中，体系的状态当用希尔伯特空间的向量来做关于体系的数学描述时被给出来. 象能量、冲量、运动的振动矩这类数量借助于自共轭运算子来研究. 例如在原子里电子的可能能级作为能量运算子的固有值而被计算. 这些固有值的差给出被原子辐射的光的频率，照这样定义了所给物质辐射的谱的构造. 电子的对应状态于此作为能量运算子固有函数而得到描述.

求量子力学问题的解常要计算不同的（常微分）运算子的固有值. 在复杂的情形下，这些问题的准确解实际上是不可能的. 为了求这些问题的近似解，广泛地使用所谓摄动理论，它使得人们能依

某自共轭运算子 A 的已知的固有值及固有函数来找与 A 有微小差异的运算子 A_1 的固有值. 注意摄动理论距完备的数学基础还远, 建立此基础是一个有趣的重要的数学问题.

另外, 与固有值的近似决定无关, 常可借定性研究的帮助讨论很多给定的问题. 在量子力学问题中, 这种研究是建立在所给的对称性问题中的有效用的基础上. 晶体对称性、原子里球对称性、关于反射的对称性等可以当做这种对称性的例子. 既然对称组成群(参见第二十章), 所以群的方法(所谓群表现论)不用计算而给出一系列问题的回答的可能性. 作为例, 我们指出原子谱分类, 核转化及其它问题.

这样, 量子力学广泛地利用自共轭运算子理论作为数学工具.

量子力学在现时的继续发展导致运算子理论进一步的发展, 在这理论前面摆着新的问题.

量子力学的影响以及泛函分析在数学内部的发展, 使得在最近若干年中代数的问题及方法在泛函分析中发生相当大的作用. 在现代分析中代数趋势的加强值得与代数方法在现代理论物理中增长着的意义(通过与十九世纪物理学方法比较)相对照.

在结束时再强调一次, 泛函分析是现代数学中正在有趣地发展着的分支之一. 它可联系并应用到现代物理学、微分方程、近似计算中, 以及代数学、拓扑学、实变数函数论中, 并由此产生了应用的一般方法, 这使泛函分析成为现代数学的基本枢纽之一.

文　　献

А. Н. Колмогоров и С. В. Фомин, Элементы теории функций и функционального анализа, вып. 1. Метрические и нормированные пространства. Издво МГУ, 1954.

Р. Курант и Д. Гильберт, Методы математической физики, т. I. Гостехиздат, 1951.

书中除其它材料外, 还陈述了振动理论与固有值理论的联系.

Л. А. 刘斯铁尔尼克, В. И. 索伯列夫, 泛函数分析概要, 科学出版社, 1955.

Ф. Рисс и Б. Секефальви-Надь, Лекции по функциональному анализу. Ил, 1954.

此书要求读者有相当的数学知识.

Г. Е. Шилов, Введение в теорию линейных пространств. Изд. 2, Гостехиздат, 1956.

阅读本书要有充分的分析学及解析几何学基础。

本章结尾涉及的问题的文献:

Б. Ван дер Варден, Метод теории групп в квантовой механике. Харьков, 1938.

И. М. Гельфанд и З. Я. Шапиро, Представления группы вращений трехмерного пространства и их применения. Успехи матем. наук, 7, №1, 3–117, 1952.

Л. О. Ландау и Е.А. Лифшиц, Квантовая механика. Гостехиздат, 1948.

田方增 译

杨宗磐 校

王建华 复校

第二十章　群及其他代数系统

§1. 引　　言

在介绍多项式代数的本书第一卷第四章中，已经谈到代数发展的基本道路、它在其他数学部门中所占的位置和关于代数对象本身的观点的改变. 本章的目的是给读者介绍某些新的代数理论的概念. 这些理论发生在上一世纪，在本世纪中得到了充分的发展. 它们对近代数学的研究有很大的影响.

近世代数，就像古典的那样，是关于运算的学说，是关于计算规则的学说. 但它不把自己局限在研究数的运算的性质上，而是企图研究更具一般性的元素上运算的性质. 这种趋向是现实中的要求所提示的. 譬如，在力学中，力、速度、旋转的合成. 在线性代数(见第十六章)中，运算的对象是矩阵，线性变换和 n 维空间的向量. 我们知道，线性代数中的想法和方法，在实际计算中被广泛的使用着.

群论在近世代数中起着特别突出的作用，本章将用很多篇幅介绍它. 在其他别的代数理论中，我们将谈到超复数系的理论. 这个理论在数的概念发展的历史过程中是一个必须的和重要的阶段. 当然，这两种理论远非穷尽近世代数的内容，但用它们可以相当清楚地阐明近世代数的想法和方法.

为了研究像对称的规律这样重要的现实世界的规律，必须找到工具. 由于这种必要性就出现了群论.

关于任何几何体或其他数学和物理对象的对称性质的知识，有时给出一个弄清这些物体、对象结构的门路. 虽然对称的概念看来是很明显的，但为了给对称这个概念一个精确的和一般的描述，特别是对称的性质的量上的计算，却需要利用群论这个工具.

群论的发生比较早，约在十八世纪末和十九世纪初．起初，它仅是作为解决这样一个问题的辅助工具而发展起来的，就是为了解决高次方程的根是否能用根号表示的问题．这是由于下面事实引起的：在解决上述问题中，第一次发现了方程的根的对称性和平等性是解决全部问题的关键．从十九世纪到二十世纪这段时间，在科学的许多别的部门，例如在几何、结晶学、物理、化学都弄清了对称的规律的重要意义．因此群论的方法和结果得到了广泛的流传．既然每一个应用部门都向群论提出自己特殊的问题，这样部门数目的增加也就必然反过来起着推进的作用，促使群论新分支的发展．而这发展把按基本概念是统一的近代群论引向分裂，分裂成为许多或多或少独立的课目：群的一般理论、有限群论、连续群论、变换的离散群、群表示与特征值的理论．并逐步地发展起来．群论的方法和概念，不仅是在研究对称的规律时，就是对于解决许多其他别的问题也是重要的了．

现在群的概念已成为现代数学中最重要的、具有概括性的概念之一了，而群论在数学中居于显著的地位．在群论的发展过程中，费得洛夫、施密特、邦特里雅金作出了杰出的贡献．就是在群论的现代发展中，苏联学者的研究也居于领导的地位．

§2．对称和变换

对称的一些最简单形式　我们先介绍一些对称的最简单形式，这些是读者从日常生活中就熟习了的．这样的对称形式之一就是几何体的镜面对称或关于平面的对称．

空间的一点 A 叫做点 B 关于平面 α 的对称点(图 1)，如果这平面垂直地交线段 AB 于其中点，通常说，点 B 是点 A 关于平面 α 的反射象．说一个几何体关于平面是对称的，如果这个平面把几何体劈成两部分，其中任一部分都是另一部分关于所给平面的镜面映象．此时这个平面就称为物体的对称平面．在自然界，镜面对称的形体非常多，例如，人体的形状，兽体和鸟体的形状通常

都有对称平面.

可以相仿地定义关于直线的对称. 通常说点 A, B 是关于直线对称的，如果该直线垂直地交线段 AB 于其中点(图 2)；说一个几何体关于直线是对称的，或者说，它以这条直线为自己的二阶对称轴，假如几何体上每一点的对称点也属于该体.

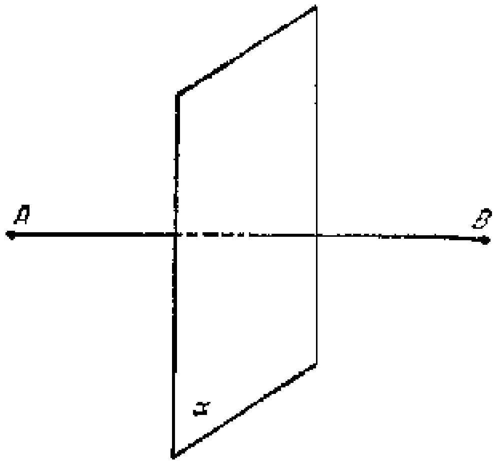


图 1

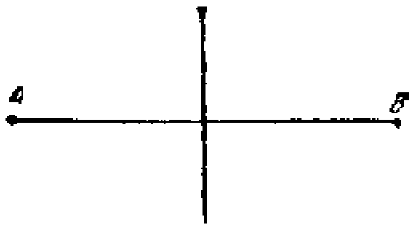


图 2

有二阶对称轴的物体，当它围绕该轴旋转一半，也就是旋转 180° 时，结果仍和自己相重.

对称轴的概念可以用很自然的方式来推广. 一直线称为给定物体的 n 阶对称轴，如果围绕轴旋转 $\frac{1}{n}360°$ 时，物体和自己相重合. 例如一个底是正 n 角形的正锥体，锥的顶点和底中心的连线是它的 n 阶对称轴(图 3).

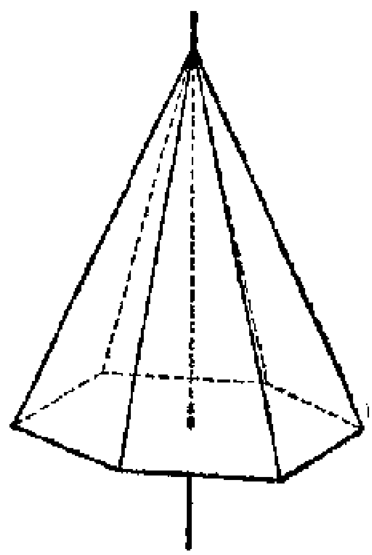


图 3

一直线称为物体的旋转轴，如果围绕着轴旋转任何角度时，物体都和自己重合. 例如，圆柱体和圆锥体的轴，球的任一条直径都是它们的旋转轴. 旋转轴同时也是任何阶的对称轴.

最后，关于点的对称，或者说中心对称，也是对称的一种重要类型. 说点 A, B 是关于中心 O 对称的，如果 A, B 所定的线段被 O 平分. 一个物体称为关于中心 O 对称，假若它的所有点

被分成一对对的、关于 O 对称的点组. 球和立方体可以作为中心对称体的例子，它们的中心就是它们的对称中心(图 4).

知道物体的所有对称平面、对称轴和对称中心，就能得到关于它的对称性质的相当全面的概念.

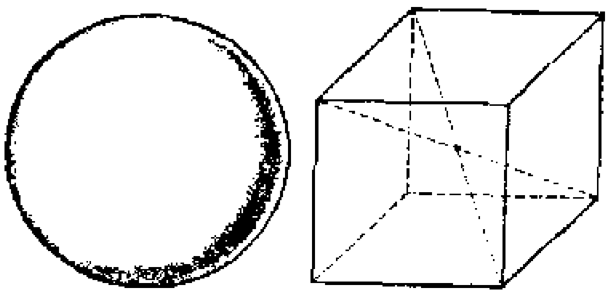


图 4

然而对称的概念不是仅当应用到几何体上才有意义. 例如下面的断语也有着完全清楚的意思: 在多项式 $x_1^3+x_2^3+x_3^3+x_4^3$ 中，变量 x_1, x_2, x_3, x_4 的出现是对称的; 而在多项式 $x_1^3+x_2^3+x_3^3+x_4^3$ 中，x_1 和 x_4, x_2 和 x_3 的出现是对称的. 但这时，例如说，变量 x_1 和 x_2 却是不同的. 这样的例子可以找到很多，这就迫使我们提出下面的重要问题: 什么是对称? 如何用数学方法来考虑对称的关系? 原来这个问题的精确答案和变换这个概念相联系着，后者从前几章开始就在本书中多次的出现了. 为了把各种不同性质的情形，例如空间物体的对称和多项式的对称，有可能包括在关于对称的一个一般定义中，必须把变换这个概念也在很一般的形式下描述.

变换 令 M 表有限或无限集合，它的元素可以是完全随意的对象. 例如，M 可以是由数 1, 2, …, n 组成的团体，可以是独立变量 x_1, x_2, x_3, x_4 的全体或平面上所有点的全体. 若对集合 M 中的每一点，在 M 中都有一个完全确定的元素和它相对应，则说给出了集合 M 的一个变换. 对于有限集合 M，每一个变换都可用表来给出. 表是由二行组成的，在上一行按任意次序记下集合 M 的元素的符号，而在它们的每一个下面注明和它对应的那个元素的符号. 例如，表 $\begin{pmatrix}1 & 2 & 3 & 4\\ 2 & 3 & 2 & 1\end{pmatrix}$ 说明集合 $\{1, 2, 3, 4\}$ 的一个变换. 在这个变换下，数 1, 2, 3, 4 顺序映到数 2, 3, 2, 1. 假若把数 1, 2, 3, 4 按次序 3, 4, 1, 2 记在上面的那一行，则我们可以把上面

谈到的变换记成表$\begin{pmatrix}3&4&1&2\\2&1&2&3\end{pmatrix}$.

假若 M 是无限集合，但是有可能排列（编号）它的元素，这时把元素的符号安置在一行（例如，M 是所有自然数的集合：1, 2, 3, …），也可用相仿的办法给出集合的变换.

为了研究变换，需要引入适当的符号. 我们将简单地用字母 A, B 等来表示变换，并且当我们用字母 A 来表示集合 M 的某一变换时，若 m 是 M 的任一元素，则 mA 就表示元素 m 的象，即指在变换 A 下 m 所对应的那个元素. 例如，令 $A=\begin{pmatrix}1&2&3&4\\2&3&2&1\end{pmatrix}$，则

$$1A=2,\ 2A=3,\ 3A=2,\ 4A=1.$$

下面指出几个在几何中起重要作用的变换.

在空间中任取一直线 a，令空间的任一点 P 与点 Q 对应，而点 Q 是点 P 绕轴 a 转一固定角度所达到的点（图 5）. 这样，我们就定义了所有空间点集合的一个变换，叫做空间绕轴 a 作角 φ 的一个旋转.

要注意的是，"旋转"这个字在力学中是指某种过程，物体上的点经这种过程而达到新的位置. 这里我们用的术语"旋转"是指空间的变换，此时不去注意运动过程的本身而是仅考察它的最终结果——点的最初位置与最终位置的对应.

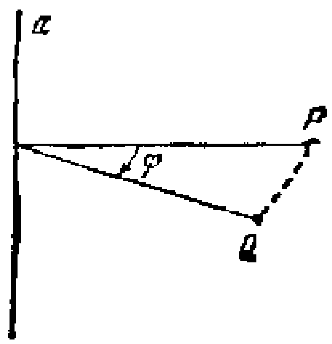


图 5

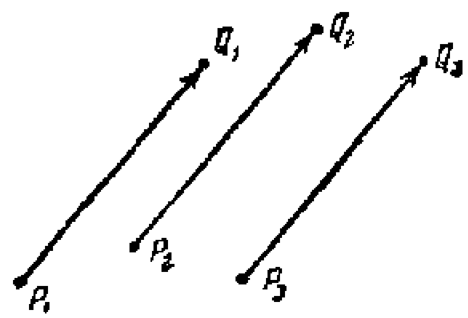


图 6

另外一种重要的空间变换是所有点按所给方向作给定距离的平移. 在图 6 中，对于任意点 P_1, P_2, P_3 示出了它们的对应点

Q_1, Q_2, Q_3. 由此可见, 假若知道在平移下某一个点的对应点, 那就可以找到空间中所有其他点的对应点.

上面曾定义过空间图形的对称平面、对称轴和对称中心等概念. 这些概念中的每一个都有空间的一个确定的变换和它对应着, 那就是关于平面的映射、关于直线的旋转和关于中心的映射. 例如, 关于平面的映射是一个变换, 在这个变换下空间的每一点与它关于平面的对称点对应着. 相仿可定义关于直线的旋转和关于中心的映射.

我们谈了一些空间的变换. 它们相应的平面的变换是: 平面绕点作给定角的旋转、平面沿自己向给定方向的平移、关于位于平面上直线的映射. 这些平面的变换可以相仿地定义, 它们较比相应的空间变换更加醒目.

双侧单值变换(即一一对应) 在考察一个集合的所有可能的变换时, 首先应注意两种映射间的基本差异, 即集合到自身的双侧单值映射及非双侧单值映射. 集合 M 的一个变换 A 叫做这个集合到自身的双侧单值映射, 假若不仅是对集合 M 的每一元素在 M 中都有唯一确定的元素和它对应——这已经包含在变换的定义中——并且对 M 的每一元素 y, 存在且只存在一个元素 x, 它所对应的元素是 y. 换句话说, 若对 M 的每一元素 y, "方程" $xA=y$ 在 M 中有一个且仅有一个解 x, 则 A 是一个双侧单值的变换.

上面考察过的所有空间变换——映射、旋转和平移——都是双侧单值的, 因为此时不仅对每一点 X 都存在一个点, X 恰好变成它, 且只存在唯一的一个点, 它能变成 X.

很容易举出反例来. 例如, 数集{1, 2, 3, 4}的由表$\begin{pmatrix}1&2&3&4\\2&1&2&3\end{pmatrix}$给出的变换就不是双侧单值的, 这是因为在该变换下没有什么数能变成数 4. 所有自然数 1, 2, 3, ⋯的集合, 由表$\begin{pmatrix}1&2&3&4&5&6&\cdots\\1&1&2&2&3&3&\cdots\end{pmatrix}$所给定的变换也不是双侧单值的. 对每一数 n 言, $2n$ 确是变到它, 但 $2n$ 不是唯一具有这性质的数, 因为 $2n-1$ 也是变到 n 的.

一般对于由表给定的变换，很容易制定一种标志，合乎它的要求时变换就是双侧单值的. 为此，易见其必要且充分条件是：在表的下一行，集合的每一元素必出现且只出现一次.

在数学里有时也讨论非双侧单值的变换. 例如，众所周知的空间到平面上投影的运算是有很大意义的. 这个变换就不是双侧单值的. 因为投影到每一个点的不是一个点，而是空间中的许多的点. 但是在大多数场合遇到的仍是双侧单值变换；特别是，当考察某些物理过程时，这样的变换起着基本的作用. 通过这些物理过程，所研究的系统的元素彼此不合并、不消失、也不产生新的.

今后谈到变换时，指的是双侧单值的变换. 它们也常被称为置换，特别是当讨论有限集合变换的情形.

对于集合 M 到自己上的每一个（双侧单值）变换 A，很容易定义一个逆变换 A^{-1}. 假若变换 A 把集合 M 的任一元素 x 变到元素 y，则把元素 y 变到 x 的变换就称为变换 A 的逆变换，用 A^{-1} 表之. 例如，若 $A=\begin{pmatrix}1&2&3&4\\2&3&4&1\end{pmatrix}$，则 $A^{-1}=\begin{pmatrix}2&3&4&1\\1&2&3&4\end{pmatrix}=\begin{pmatrix}1&2&3&4\\4&1&2&3\end{pmatrix}$；若 A 是空间绕轴作角 φ 的旋转，则 A^{-1} 是围绕同一个轴，但向相反方向作角 φ 的旋转；等等.

有时可能碰到逆变换和原给的相重合的情形. 比如说，在空间中关于平面或点的映射就有这样的性质. 置换 $A=\begin{pmatrix}2&1&4&3\\1&2&3&4\end{pmatrix}$ 也有此性质，因为 $A^{-1}=\begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix}=\begin{pmatrix}2&1&4&3\\1&2&3&4\end{pmatrix}$.

注意，对于非双侧单值变换不能谈什么逆变换，因为这样的变换可能把谁也变不到某些元素上去，另一方面也可能把同一元素变到许多元素上去.

对称的一般定义 在数学和它的应用中，很少会发生考虑给定集合的一切变换的要求，这是因为集合本身很少能被认为仅仅是一些互不相关的元素的简单联合. 这也是很自然的，因为在数学里研究的集合是许多具体集合的抽象形象，而这些具体集合的

元素是永远处于千千万万的相互联系中的，且与所考察的集合范围以外的一些东西相联系着．显然在数学中不得不撇开这些联系的大部分，但最本质的部分是要保留下来并加以考虑的．这就促使我们首先研究集合的这样一些变换，它们不破坏元素间这些或那些被考虑的联系，通常称这样的变换为可容许变换或称为对集合元素间被考虑的那些联系的自同构对应．例如，对于空间的点，两点间距离的概念是很重要的．这个概念的存在是考虑到下面的这个点间的联系：任两点间都有一定距离．不破坏这些联系的变换是使点间距离不变的变换．这样的变换叫做空间的“运动”．

利用自同构这个概念，不难给出对称的一般定义．设给定一集合 M，在其内考虑元素间的某些关系，并设 P 是 M 的一个子集．对于 M 的一个可容许变换 A，说集合 P 是对称的或不变的，若变换 A 把集合 P 中的每一点仍变为 P 的点．因此集合 M 的把 P 仍变到自己的可容许变换的全体就刻划了集合 P 的对称．我们的定义完全适用于在空间内物体对称的概念．把整个空间看做集合 M，把“运动”看做是可容许变换，把给定的物体看做 P，这样，使 P 仍回到自己的运动的全体就刻划了体 P 的对称．

从前讨论过的空间的映射、平移和绕给定直线的旋转是运动的一些特殊情形，因为在这些变换下点间的距离显然是不变的．更详细的研究就会知道，平面的每一运动或是平移，或是绕中心的旋转，或是关于直线的映射，或是关于直线的映射和顺着这条直线的平移的联合．相仿地，空间的每一运动或是平移，或是绕轴的旋转，或是扭转运动，即伴有着沿旋转轴平移的绕轴旋转，或是关于平面的映射，也可能还伴有一个沿映射平面的平移或是一个绕垂直于该平面的轴的旋转．

空间的平移、旋转和扭转称为它的固有运动或第一类运动．其他的“运动”(包括映射在内)叫做非固有运动或第二类运动．在平面上第一类运动是平移和旋转，而关于直线的映射以及伴有旋转或平移的映射被认为是第二类运动．

容易想象，凡第一类运动的变换可以看做是空间到自身的或

平面到自身的连续运动的结果. 这样去看第二类的运动是不可能的,因为作为它们的组成部分的映射破坏了这一点.

常常说平面在自己所有部分上都是对称的, 或是说平面的点都是平等的. 用变换这个精确的语言来说明这断语, 那就是借助一个适当挑选的"运动"可以使平面任意一点与任意另外的一个点相重合.

从前考察过的一些物体或图形对称的情形也是包含在对称的一般定义内的. 例如,关于平面 α 对称的物体,在关于平面 α 的映射下是回到自身的; 关于中心 O 对称的物体, 在关于 O 的映射下是回到自身的. 因此空间的使一物体或一空间图象和自己重合的第一类和第二类运动的全体, 就充分刻划了这个物体或空间图象对称性的程度. 上面指出的这个运动的集合愈丰富,种样愈多,则该物体或图象对称性的程度也就愈大. 特别是, 假若这个集合除去恒等变换外不包括任何运动, 则可以称该物体或图象为不对称的.

使平面上的正方形仍回到自身的平面运动的全体刻划了正方形对称性的程度. 但若正方形和自己重合, 则两对角线的交点也必然要和自己重合. 因此,所求运动使正方形中心不动,因而或是绕中心的旋转, 或者关于通过中心的直线的映射. 由图 7 容易看到, 正方形 $ABCD$ 是关于绕其中心作 90° 倍数角的旋转对称的, 对于关于对角线 AC, BD 和直线 KL, MN 的映射, 它也是对称的. 这八个运动也就刻划着正方形的对称.

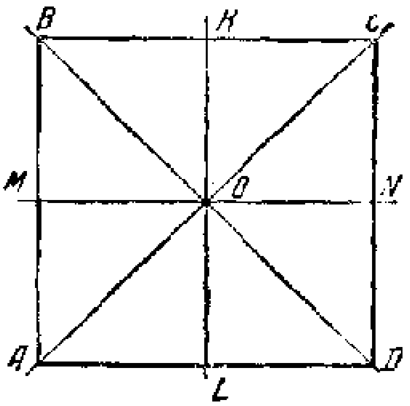


图 7

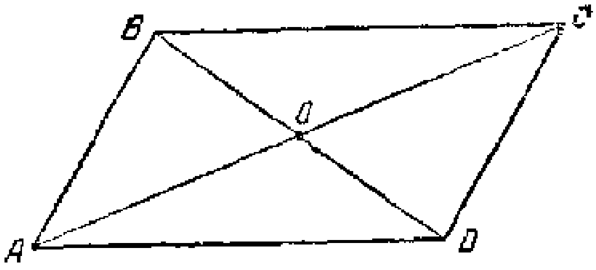


图 8

长方形的对称的集合由绕中心作 180° 的旋转和关于对边中点连线的映射组成，而平行四边形（图 8）对称的集合则仅由绕中心作 180° 倍数角的旋转组成，即由关于中心的映射和恒等变换组成.

上面我们引用了代数里的对称的例子，即曾指出，多变数多项式对称的概念是有意义的.

现在来考察如何刻划多项式的对称．我们说，在多项式 $F(x_1, x_2, \cdots, x_n)$ 内完成了未知量的一个置换 $A=\begin{pmatrix} x_1, & x_2, & \cdots, & x_n \\ x_{i_1}, & x_{i_2}, & \cdots, & x_{i_n} \end{pmatrix}$ 或简写成 $A=\begin{pmatrix} 1, & 2, & \cdots, & n \\ i_1, & i_2, & \cdots, & i_n \end{pmatrix}$，是指在这个多项式内所有的地方都把字母 x_1 换成 x_{i_1}，x_2 换成 x_{i_2}，……等等．象这样得到的多项式用 FA 表示．例如，若 $F=x_1^2-2x_2+x_3-x_4$，$A=\begin{pmatrix} 1\,2\,3\,4 \\ 3\,1\,4\,2 \end{pmatrix}$，则 $FA=x_3^2-2x_1+x_4-x_2$.

给定的多项式的对称，是由下面那些未知量的置换所作成的集合来刻划，当在多项式上完成了这些置换，它们使该多项式不变．例如，多项式 $x_1^2+2x_2+x_3^2+2x_4$ 的对称可由下面四个置换来刻划：$\begin{pmatrix} 1\,2\,3\,4 \\ 1\,2\,3\,4 \end{pmatrix}$，$\begin{pmatrix} 1\,2\,3\,4 \\ 3\,2\,1\,4 \end{pmatrix}$，$\begin{pmatrix} 1\,2\,3\,4 \\ 1\,4\,3\,2 \end{pmatrix}$，$\begin{pmatrix} 1\,2\,3\,4 \\ 3\,4\,1\,2 \end{pmatrix}$，而多项式 $x_1^2+2x_2+x_3^2+x_4$ 的对称仅由两个置换来刻划：$\begin{pmatrix} 1\,2\,3\,4 \\ 1\,2\,3\,4 \end{pmatrix}$，$\begin{pmatrix} 1\,2\,3\,4 \\ 3\,2\,1\,4 \end{pmatrix}$.

§3. 变 换 群

变换的乘法 在研究变换的性质时，很容易发现，某些变换可由另外一些变换组成．例如，螺旋运动由绕轴的旋转及延轴的移动组成．这个由给定的变换组成新变换的过程叫做变换的乘法．我们对集合 M 的一个任意元素 x 进行变换 A，然后对新元素 xA 运用变换 B，这样就得到元素 $(xA)B$．把 x 直接变到 $(xA)B$ 的变

换称为变换 A 和 B 的乘积，用 AB 表示．因此，按照定义，有

$$x(AB)=(xA)B.$$

例如

$$\begin{pmatrix}1&2&3&4\\2&3&4&1\end{pmatrix}\begin{pmatrix}1&2&3&4\\3&4&1&2\end{pmatrix}=\begin{pmatrix}1&2&3&4\\4&1&2&3\end{pmatrix}.$$

这是因为第一个置换将1变到2，第二个置换将2变到4，因而它们的乘积应把1变到4，等等．再看几个例子：

$$\begin{pmatrix}1&2&3&4\\3&1&4&2\end{pmatrix}\begin{pmatrix}1&2&3&4\\1&3&2&4\end{pmatrix}^{-1}=\begin{pmatrix}1&2&3&4\\3&1&4&2\end{pmatrix}\begin{pmatrix}1&2&3&4\\1&3&2&4\end{pmatrix}=\begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix};$$

$$\begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix}\begin{pmatrix}1&2&3&4\\3&1&2&4\end{pmatrix}=\begin{pmatrix}1&2&3&4\\1&3&4&2\end{pmatrix};$$

$$\begin{pmatrix}1&2&3&4\\3&1&2&4\end{pmatrix}\begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix}=\begin{pmatrix}1&2&3&4\\4&2&1&3\end{pmatrix}.$$

最后两个例子说明，变换的乘法是不可交换运算，它的结果和因子的顺序有关．对于平面上运动的乘法，这样事实也是容易肯定的．例如，A 是平面上绕原点作 $90°$ 的旋转，而 B 是延 Ox 轴平移一个单位长．

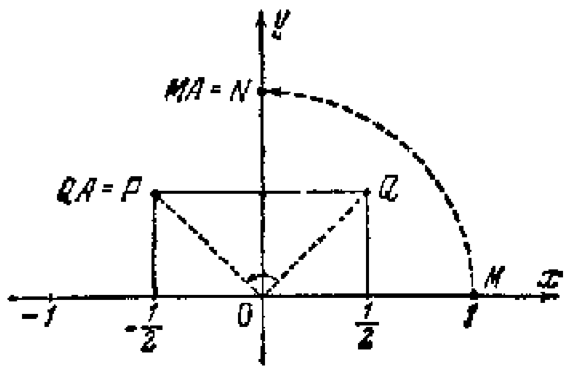


图 9

今考察，变换 AB 和 BA 把原点 O 变为哪个点．按定义，有(图 9)

$$O(AB)=(OA)B=OB=M,$$
$$O(BA)=(OB)A=MA=N,$$

即 $AB\neq BA$.

为了弄清楚变换 BA 的几何性质，我们来考察点 P. 我们有

$$P(BA)=(PB)A=QA=P,$$

即点 P 在变换 BA 下停留不动．从这一点容易判断，BA 就是平面上一个绕点 P 做 $90°$ 的旋转．类似地

$$Q(AB)=(QA)B=PB=Q,$$

而 AB 是平面上一个绕点 Q 做 $90°$ 的旋转．

平面或空间中运动的乘法规则，在一般情形是相当复杂的。但在两种重要的情形下，乘法的规则是很简单的．首先，若平面上绕同一点作角 φ 和角 ψ 的旋转相乘，或空间内绕同一直线作角 φ 和角 ψ 的旋转相乘，则其结果将是相应的作角 $\varphi+\psi$ 的旋转．其次，若两个顺序被向量 $\overrightarrow{MN}$ 和 $\overrightarrow{NP}$ 所刻画的平移相乘，则其乘积仍是平移，它由向量 $\overrightarrow{MP}$，即最初的那两个向量之和，所刻划。

变换"乘法"这个名词本身意味着在数的乘法和变换的乘法之间有某种类似，但是这种类似并不是完全的．譬如说，对数的乘法有可换(交换)律．我们已看到，在进行变换的乘法时，这个律可能被破坏．算术的第二个基本律——乘法的结合律——对于变换的乘法完全适合，即对集合 M 的任意三个变换 A, B, C, 都有 $A(BC)=(AB)C$。

事实上，若 m 是 M 中任意元素，则

$$m[A(BC)]=(mA)(BC)=[(mA)B]C$$
$$=[m(AB)]C=m[(AB)C].$$

结合律容许用乘积 $A(BC)=(AB)C=ABC$ 代替变换 A, B, C 的两个乘积 $A(BC)$ 及 $(AB)C$．结合律还说明，四个或更多个变换的乘积和括号的分布无关．

变换中有一个变换，它相当于数 1，就是使 M 的每一个元素不动的恒等变换，或称单位变换 E．显而易见 $AE=EA=A$，不论 A 是什么样的变换．

我们还要指出下面重要事实：双侧单值变换的乘积是双侧单值的．事实上，为了在集合 M 中找到一元素 x，使它在变换 AB 之下的对象恰是给定的元素 a，只须先找到一个被变换 B 变为元素 a 的元素 x_1，然后再找到一个被变换 A 变为元素 x_1 的元素 x_2。因为 $x_2(AB)=(x_2A)B=x_1B=a$，故 x_2 就是想找的元素 x。

变换 A 及其逆变换 A^{-1} 的乘积是单位变换，即

$$AA^{-1}=A^{-1}A=E.$$

这个直接由逆变换的定义得出。

上面考虑的平面的平移和旋转相乘的例子说明，从乘积中因子的性质，未必总能看出变换乘积的性质．但是形如 $C=B^{-1}AB$ 的变换乘积是一个重要的例外．C 的性质在这里非常简单的和 A, B 的性质相联系着，即，若集合 M 中的元素 m 被变换 A 变为元素 n，则被变换 B"移动"了的 mB 元素被变换 C 变为被"移动"了的元素 nB．

证明：$(mB)B^{-1}AB=mAB=nB$．

变换 $B^{-1}AB$ 称为用 B 去改变 A 的方法由 A 得来的，或称为借助于 B 与 A 相共轭的变换．

例如，我们用平移来改变绕点 O 平面旋转 P_0．根据引用过的规则，为了找对改变了的运动 $C=V^{-1}P_0V$ 而言点的开始位置和最终位置的对偶，应当利用 V 去移动对变换 P_0 而言的点的相应对偶．因为点 O 在旋转 P_0 下是和自己成对的（图 10），故点 OV 对于变换 C 是和自己成对的．还有，假若点 M 被变换 P 变为点 N，则被移动了的点 MV 将被变换 C 变为点 NV．因此从图 10 可看到，变换 C 是绕点 OV 作与旋转 P_0 相同角度 φ 的旋转．

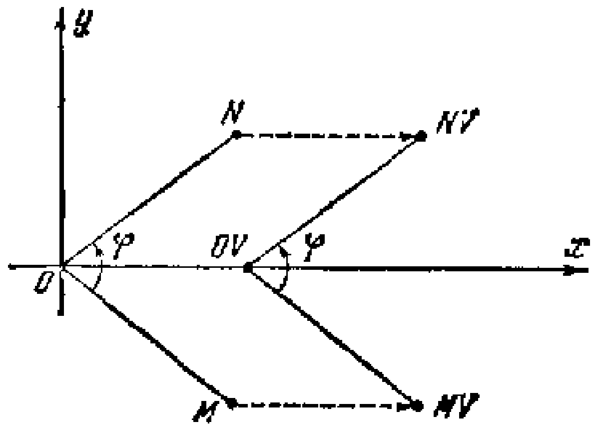


图 10

可以类似证明，假若借助于作角 φ 的旋转 P_0 去改变平面上的平移，而后者是用向量 $\overrightarrow{MN}$ 来刻划的，则仍将得到平面上的平移，而它是被已旋转过的向量来刻划的．

即使在变换是由表给出的情形，上面所说求变换 $B^{-1}AB$ 的法则，也能非常漂亮地叙述出来．令

$$A=\begin{pmatrix}1 & 2 & \cdots & n\\ a_1 & a_2 & \cdots & a_n\end{pmatrix},\quad B=\begin{pmatrix}1 & 2 & \cdots & n\\ b_1 & b_2 & \cdots & b_n\end{pmatrix}$$

此时有

$$B^{-1}AB=\begin{pmatrix} b_1\ b_2\cdots b_n \\ 1\ \ 2\cdots n\end{pmatrix}\begin{pmatrix}1\ \ 2\ \cdots\ n \\ a_1\ a_2\cdots a_n\end{pmatrix}\begin{pmatrix}1\ \ 2\cdots n \\ b_1\ b_2\cdots b_n\end{pmatrix}=\begin{pmatrix}b_1\ b_2\ \cdots\ b_n \\ b_{a_1}\ b_{a_2}\cdots b_{a_n}\end{pmatrix},$$

即，为了利用置换 B 来改变置换 A，只需把置换 A 的上列和下列中的数字用置换 B 所决定的变换作用一下．例如，若

$$A=\begin{pmatrix}1\,2\,3\,4\,5 \\ 3\,5\,4\,1\,2\end{pmatrix},\ B=\begin{pmatrix}1\,2\,3\,4\,5 \\ 2\,5\,1\,3\,4\end{pmatrix},$$

则

$$B^{-1}AB=\begin{pmatrix}1B2B3B4B5B \\ 3B5B4B1B2B\end{pmatrix}=\begin{pmatrix}2\,5\,1\,3\,4 \\ 1\,4\,3\,2\,5\end{pmatrix}=\begin{pmatrix}1\,2\,3\,4\,5 \\ 3\,1\,2\,5\,4\end{pmatrix}.$$

在这里提一下，虽然在一般情形下两个变换的乘积和它们的顺序有关，但在个别的情况，乘积 AB 和 BA 是可能相同的．此时变换 A 和 B 称为可换位的或可换的．若 $AB=BA$，则

$$B^{-1}AB=B^{-1}BA=A.$$

因此，用一个与给定置换可换的置换去改变它，将不影响这个给定的置换．

变换群 能刻划某一形体的对称的所有变换所组成的集合并不是随意的，这样的集合显然具有下列性质：

1．集合中任两变换的乘积仍属于这个集合；

2．恒等变换属于这样的集合；

3．若一变换属于这样的集合，则其逆变换也属于该集合．

这些性质对于研究变换是非常重要的．由一集合 M 的双侧单值变换所组成的且具有上面列举的三个性质的集合叫做集合 M 的变换群，不论这个变换的集合是否刻划某一形体的对称．

从代数学的观点来看，性质 1—3 是很基本的，因为它们使得人们能够由给定的集合中的一些变换 A, B, C, …组成各种各样形状如 $ABAC$, $A^{-1}BCB^{-1}$ 等的新变换，并且性质 1—3 保证了这样得到的所有变换，都不出给定的变换集合的圈子．

一个群所含变换的数目称之为群的元素；元素可能是有限的或无限的．和这个相应的，群分成有限的无限的．上面曾讨论过平面上正方形的对称群，这个群包含八个变换；另一方面，平面上

点 A_i 所组成的无限集合，如图 11 所示，被下列平面上的运动变为自身：顺着 OA 轴向左或向右移动 OA 长的倍数；关于虚线的反射；关于 OA 轴的反射．由此可见，这个形体的对称群是无限的．

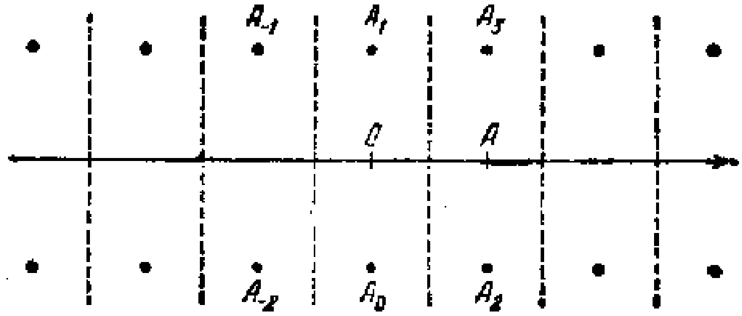


图 11

保持某一物体，也就是说，刻划它的对称的变换的全体永远是一个群．用对称群的形式来给出群的这个方法是非常重要的方法之一．按着这个原则可得到最重要的群．它们中首先要举出的是平面和空间的运动群．正多面体的对称群也很值得重视．如众所周知的，在空间中仅有五种正多面体(4，6，8，12 和 20 面体)．任取正多面体中的一个并考察把它仍变回自身的所有空间运动，我们就得到群——这个多面体的对称群．假若代替所有的运动只考察把多面体变回自身的第一类运动，则所得到的还是一个群，这个群是整个的多面体对称群的一部分，称此群为多面体的旋转群．因为当多面体和自身相重合时，其中心也必和自己相重合，所以所有属于多面体对称群的运动都使多面体的中心保持不动，因此它们只能或者是绕通过中心的轴的旋转，或者是关于通过中心的平面的映射，或者，最后一种可能，是伴有绕轴旋转的关于上述那样平面的映射，而这里的旋转轴是通过中心且垂直于这些平面．

利用这些事实，容易找到所有正多面体的对称群和旋转群．在表 1 中示出正多面体的对称群和旋转群的元数．所有这些群都是有限的．

置换群 在数学里，历史上第一个被研究的变换群是 n 元多项式中变量 $x_1, x_2, \cdots, x_n$ 的置换群．这样群的研讨和高阶方程式

表 1

面　　数	4	6	8	12	20
对称群的元数	24	48	48	120	120
旋转群的元数	12	24	24	60	60

的解可否用根式来表示的问题紧密联系着. 显然, 具下面性质的某些变量的置换的全体作成一群: 它们使含这些变量的一个或几个多项式的值不变. 在变量的所有置换下都不改变的多项式叫做对称多项式. 例如, $x_1+x_2+\cdots+x_n$ 是对称多项式. 相应地, 某一给定的变量集合的所有置换的全体称为这个集合的对称置换群.

被摆置的变量的数目称为对称群的阶. 代替变量 $x_1, x_2, \cdots, x_n$ 的置换, 可以考察数字 1, 2, $\cdots$, n 的置换. 由于数字的每一置换都能写成$\begin{pmatrix}1 & 2 & \cdots & n\\ a_1 & a_2 & \cdots & a_n\end{pmatrix}$的样子, 其中 $a_1, a_2, \cdots, a_n$ 是数字 1, 2, $\cdots$, n 按某种顺序写成的, 故所有置换的总数等于 n 个元素的排列总数, 即对称群的元数等于 $n!=1\cdot2\cdot3\cdots n$. 这个元数是随着 n 的增大而迅速增长的, 当 $n=10$ 时, 对称群的元数为 3628800.

我们考虑多项式

$$F(x_1, \cdots, x_n)=(x_2-x_1)(x_3-x_1)\cdots(x_n-x_1)(x_3-x_2)\cdots (x_n-x_2)\cdots(x_n-x_{n-1}). \tag{1}$$

显然, 每一个置换或者不改变多项式 F 的值, 或者仅改变 F 的符号. 第一种置换叫做偶置换, 改变 F 符号的置换叫奇置换. 全体偶置换的集合组成多项式(1)的对称群, 称这个群为交代置换群.

任意两个偶置换的乘积是偶置换, 因此全体偶置换组成群. 两个奇置换的乘积是偶置换.

事实上, 如果 A, B 是两个奇置换, 则

$$FAB=(FA)B=(-F)B=-(-F)=F.$$

同样方法可以证明, 偶置换与奇置换的乘积是奇置换. 偶置换的逆或者奇置换的逆与本身有同样奇偶性.

仅交换元素1和2位置的置换$\begin{pmatrix}1 & 2 & 3 & \cdots & n\\ 2 & 1 & 3 & \cdots & n\end{pmatrix}$,可以作为奇置换的例子.

置换分解成循回的乘积 在研究置换群时，将置换表成所谓循回的乘积有着巨大的意义. 首先,定义符号$(m_1, m_2, \cdots, m_k)$为这样的置换，它把m_1变到m_2，把m_2变到m_3，…，把m_{k-1}变到m_k，最后，又把m_k变到m_1，而所讨论集合中所有其余的元素均不变. 例如,如果讨论数字1, 2, 3, 4, 5的置换,那么

$$(1, 2, 3, 4, 5)=\begin{pmatrix}1 & 2 & 3 & 4 & 5\\ 2 & 3 & 4 & 5 & 1\end{pmatrix},\ (3, 5)=\begin{pmatrix}1 & 2 & 3 & 4 & 5\\ 1 & 2 & 5 & 4 & 3\end{pmatrix}.$$

置换$(m_1, m_2, \cdots, m_k)$叫做长为k的循回；$m_1, m_2, \cdots, m_k$叫做循回的元素. 单位置换约定写成长为1的循回$(1)=(2)=\cdots$的形式. 长为2的循回叫做一个对换. 在循回中循环推移循回的元素，我们仍得到循回本身，例如$(1, 2, 3)=(2, 3, 1)=(3, 1, 2)$；$(5, 6)=(6, 5)$.

容易验证，没有共同元素的两个循回，例如$(2, 3)$，$(1, 4, 5)$是可换的置换. 因此,在这样的循回相乘时,可以不考虑乘积中因子的顺序.

在一般理论中，循回的意义基于下面的定理：每一个置换可以表成无共同元素的循回的乘积,而且除因子顺序外,表法是唯一的.

定理的证明由下面的表示法直接可以看出. 假定我们希望分解置换$A=\begin{pmatrix}1 & 2 & 3 & 4 & 5 & 6\\ 4 & 5 & 6 & 3 & 2 & 1\end{pmatrix}$. 我们看到，$A$把1变到4，4变到3，3变到6，6变到1. 这样，我们便得到了要表成的乘积的第一个因子$(1, 4, 3, 6)$. 还余下两个数目，我们看到，A把2变到5，5变到2. 因此，第二个因子是$(2, 5)$. 由于所有元素都被讨论到了，故

$$\begin{pmatrix}1 & 2 & 3 & 4 & 5 & 6\\ 4 & 5 & 6 & 3 & 2 & 1\end{pmatrix}=(1, 4, 3, 6)(2, 5). \tag{2}$$

将置换分解为有共同元素的循回的乘积也是可能的，但这样的分解不是唯一的．例如，

$$(a_1, a_2, \cdots, a_n)=(a_1, a_2)(a_1, a_3)\cdots(a_1, a_n)$$
$$=(a_2, a_3)(a_2, a_4)\cdots(a_2, a_n)(a_2, a_1). \qquad (3)$$

我们来证明，每一个长为 2 的循回是奇置换．我们已经知道，对于$(1, 2)$的循回是正确的．但任一循回(i, j)等于$S^{-1}(1,2)S$，此处 S 是把 1 变到 i，2 变到 j 的任意置换$\begin{pmatrix}1 & 2 & \cdots \\ i & j & \cdots\end{pmatrix}$．因为$(1, 2)$是奇置换，而 S 与 S^{-1} 的奇偶性相同，故 $S^{-1}(1, 2)S$ 是奇置换．

按照公式 (3)，长为 $m+1$ 的循回可以表成 m 个奇置换的乘积．因之，如果 $m+1$ 是奇数，则长为 $m+1$ 的循回是偶置换，如果 $m+1$ 是偶数，则长为 $m+1$ 的循回是奇置换．这样，如果置换分解成循回的乘积是已知的，那么，便可以很快的计算出这个置换的奇偶性．例如，置换$\begin{pmatrix}1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 3 & 2 & 1\end{pmatrix}$是偶置换，因为按照(2)，它是两个偶置换的乘积．

子群 若群的一部分也作成群，则称之为原来群的一个子群．这样，变量 $x_1, x_2, \cdots, x_n$ 的置换的交代群是这些变量的置换的对称群的一个子群．平面的固有运动的全体组成一个群，这个群是平面的固有运动与非固有运动的全体所组成的群的一个子群．

从纯形式的观点看来，单位(恒等)变换本身也组成一个子群．同样，任意群本身也可视为它自己的子群．然而，除了这两个当然的子群以外，几乎每一个群均含有好多子群．知道群的所有子群的意义在于它给出关于已知群内部结构的足够完备的表示．

构成子群的一个相当广泛的方法，是指出所谓子群的生成元的方法．

设 $A_1, A_2, \cdots, A_m$ 是属于群 G 的某些变换．作这些变换以及其逆的一切可能乘积的集合 H，则 H 是一个群．事实上，单位变换属于 H，因其可表成 $A_1A_1^{-1}$ 的形式．其次，如果 B, C 可表

成上述乘积的形式，那末 BC 也能表成上述乘积，故属于 H. 最后，如果 B 能表成上述乘积，例如，$B=A_1^{-1}A_2A_1A_1A_2^{-1}$，则 B^{-1} 也能表成所要求乘积的形式，即 $B^{-1}=A_2A_1^{-1}A_1^{-1}A_2^{-1}A_1$.

群 H 显然是群 G 的子群，称之为由 $A_1, \cdots, A_m$ 生成的子群，而变换 $A_1, A_2, \cdots, A_m$ 本身，叫做子群 H 的生成元. H 与 G 可能重合，这时，称 $A_1, A_2, \cdots, A_m$ 为群 G 本身的生成元. 不难由例子看出，同一子群可以有不同的生成元组.

由一个变换 A 生成的子群叫做循环群，它的元素是变换

$$E, A, AA, AAA, \cdots, A^{-1}, A^{-1}A^{-1}, A^{-1}A^{-1}A^{-1}, \cdots.$$

很自然的称其为变换 A 的方幂，即

$E=A^0$, $A=A^1$, $AA=A^2$, $\cdots$, $A^{-1}A^{-1}=A^{-2}$, $A^{-1}A^{-1}A^{-1}=A^{-3}$, $\cdots$.

和在普通算术中一样，容易证明

$$A^mA^n=A^{m+n} \text{ 以及 } (A^m)^n=A^{mn}. \tag{4}$$

变换说是**循环**的，如果它的某个方幂等于恒等变换. 使循环变换成为恒等变换的最小正指数叫做**变换的阶数**(或周期). 我们约定非循环变换的阶数是无限的.

我们来讨论几个例子. 设 A 是平面绕 O 点所作 $\frac{360^\circ}{n}$ 的旋转，此处 n 是大于 1 的正整数. 于是 A^2 是旋转 $2\frac{360^\circ}{n}$ 角的旋转; A^3 是旋转 $3\frac{360^\circ}{n}$ 的旋转; A^{n-1} 是旋转 $(n-1)\frac{360^\circ}{n}$ 角的旋转; 而 A^n 是作 360° 的旋转，即恒等变换. 这就表明平面绕 O 点所作 $\frac{360^\circ}{n}$ 的旋转是 n 阶循环变换.

设 A 是平面延某个直线所作的平移，于是 A^2, A^3, $\cdots$也是平面延同一直线的平移，但移动的距离分别是原来所移动的距离的二倍、三倍等等. 因之，A 的任何正方幂也不能是恒等变换，即 A 的阶数是无限的.

元素 A 所生成的循环群的元素是

$$\cdots A^{-2},\ A^{-1},\ E,\ A,\ A^2,\ \cdots. \tag{5}$$

若 A 是无限阶变换，则序列(5)中的所有变换都是互异的，因之，这个循环群是无限群．事实上，如果不是这样，我们将得出等式 $A^k=A^l(k<l)$．由此可知 $A^{l-k}=E,\ l-k>0$，这与 A 不是循环变换矛盾．

现在设 A 是 m 阶循环变换，于是

$A^m=E,\ A^{m+1}=A,\ A^{m+2}=A^2,\ \cdots,\ A^{m-1}=A^{-1},\ A^{m-2}=A^{-2}$，$\cdots$，即序列(5)由变换 $E,\ A,\ A^2,\ \cdots,\ A^{m-1}$ 所组成，并且这 m 个变换中任意两个都不相等，因为，如果有 $A^k=A^l(0\leqslant k<l<m)$，则将有 $A^{l-k}=E,\ 0<l-k<m$，这与 A 的阶数为 m 矛盾．因之，阶数为 m 的循环变换所生成的循环群，由 m 个不同的变换所组成．

所有元素彼此间都可换的群叫做可换群或者阿贝尔群，是以挪威数学家阿贝尔的名字来命名的，他发现了这样的群对于方程论的重要意义．

式子(4)表明，同一变换的方幂彼此间永远是可换的：$A^mA^n=A^nA^m=A^{m+n}$．因之，循环子群恒为阿贝尔群．

在数的算术中，除法同乘法起着重要作用．在群论中，由于乘法的不可交换性，必须谈到两种除法：左除和右除．事实上，解方程 $Ax=B$，此处 $A,\ B$ 是给定的变换，而 x 是要求的变换，自然地引出用 A 除 B 的左商；而解方程 $yA=B$，引出用 A 除 B 的右商．在第一个等式中，两端同用 A^{-1} 左乘，而在第二个等式中，两端同用 A^{-1} 右乘，我们得到 $x=A^{-1}B,\ y=BA^{-1}$．因此，可以把 $A^{-1}B$ 或者 BA^{-1} 看作为变换 B 被 A 除的"商"．

从很多例子看到，在群里通常 $AB\neq BA$．"商"$(AB)\cdot(BA)^{-1}$ 或者 $(BA)^{-1}(AB)$ 可以采取作为置换 A 与 B 的不可交换性的"测度"．上面第二个表示式 $(BA)^{-1}(AB)=A^{-1}B^{-1}AB$ 称为 A 与 B 的换位子，并且用符号 $(A,\ B)$ 表示．由式子

$$(A,\ B)=A^{-1}B^{-1}AB$$

可以看出，换位子可以表为共轭变换 $B^{-1}AB$ 被第一个变换 A 除的"商"．

例如，设 A 是平面的平移，则共轭变换也将是平移，而两个平移的商显然是一个平移. 因之，平移与任意变换的换位子是一个平移. 设 A 是绕某个定点 O 作 φ 角的旋转，而 B 是旋转或者平移. 于是，共轭变换仍是旋转 φ 角的旋转，然而是绕另外一个点 O'. 因之，在这个情形，换位子(A, B)是绕 O 点作负角 φ 的旋转与绕 O' 点作正角 φ 的旋转的乘积. 由图 12 可以看出，最后得出的变换是沿着与线段$\overline{OO'}$成$\left(\frac{\pi}{2}-\frac{\varphi}{2}\right)$角的直线作距离为 $2\cdot\overline{OO'}\sin\frac{\varphi}{2}$ 的平移.

因此，我们必须注意这样有趣的事实：平面上任意两个第一类运动的换位子是平移或恒等变换. 由于 $(A, B)=E$ 表示 $AB=BA$，故在平面上第一类运动所作成的任意一个非交换群均含有平移.

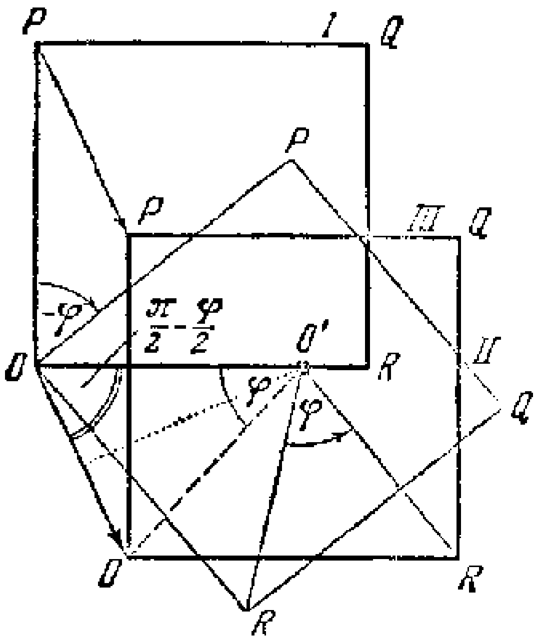


图 12

群 G 中所有元素的换位子生成的子群叫做 G 的**换位群**，换言之，群 G 的换位群由一切可表成换位子乘积的元素所组成. 由于平面上任意两个第一类运动的换位子是平移，而平移的乘积仍是平移，故平面上第一类运动群的换位群仅由平移所组成.

阿贝尔群的换位群仅由一个单位变换所组成，因为由 $AB=BA$ 可得出$(A, B)=E$.

设 G 是数字 1, 2, …, n 的所有置换作成的对称群. 我们来证明，任意两个置换 A, B 的换位子永远是偶置换. 事实上，置换 AB, BA 因之$(BA)^{-1}$都有相同的奇偶性，于是，作为有同一奇偶性的两个置换$(BA)^{-1}$, (AB)的积(A, B)是偶置换.

我们看到，对称群的换位群仅含有偶置换. 容易证明，它与交

代群重合.

群 G 的换位群也叫做 G 的导出群, 用符号 G' 表示. G' 的换位群叫做 G 的二次换位群, 用符号 G'' 表示. 把这个过程重复下去, 我们得到 G 的任意次换位群.

G 的各次换位群中, 如果有一个仅由单位变换所组成(于是, 这个换位群后面的各次换位群均由单位变换组成), 那么, 就说 G 是可解群. 这个名称起源于方程论, 在这里可解群与方程可用根号解相对应. 平面上第一类运动群是可解群, 因其第二次换位群等于单位元. 二次、三次、四次对称群也是可解群, 因为它们的一次、二次、三次换位群分别等于单位元. 然而, 五次以及五次以上对称群不是可解群, 因为可以证明, 它们的二次换位群与一次换位群重合, 而一次换位群是不等于单位元的.

§4. 费得洛夫群

有限平面形的对称群 就像已经说过的一样, 形或体的对称性可由平面或空间中使形体仍和自己重合的运动群来刻划.

平面上有限形[1]的对称群是最容易被找到的. 事实上, 设给出平面上任一有限形, 并设此形经某一运动 A 后和自己重合. 此时形之重心 O 经运动 A 必仍和自己重合, 即 A 或是绕 O 点的旋转或是关于过 O 点直线的翻褶. 这样, 任意有限平面形的对称群仅由绕它的重心的旋转及关于过重心的直线的翻褶组成.

我们将顺序考察在研究有限形的对称群时可能出现的不同情形.

1. 仅由单位(恒等)变换所组成的对称群 K_1. 这是任意非对称形(图 13)的对称群.

2. 由单位变换及关于某一直线的翻褶组成的对称群(图 14).

今指出, 若群 K 含有关于过 O 点及夹角为 α 的两直线的翻

1) 有限性了解为: 整个图形位于平面中有限部分, 例如说在某个圆内.

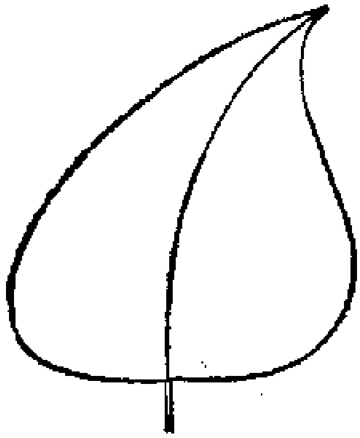


图 13

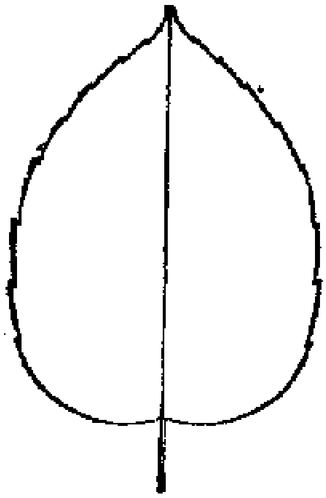


图 14

褶，则此两翻褶的积是绕 O 点作 2φ 角的旋转(图 15)．由之可见，群 K_2 是唯一不含旋转的对称群．

3．只由一些旋转组成的对称群，但其中不含作任意小角的旋转．在此种情况，在 K_3 的旋转可找到一个作最小正角的旋转．设此角为 α^0．今证，群中的旋转所作之旋转角必是 α^0 之倍角．设 β^0 为某旋转的角的度数，可找到一整数 h，使 $h\alpha^0\leqslant\beta^0<(h+1)\alpha^0$，由之 $0\leqslant\beta^0-h\alpha^0<\alpha^0$．群 K_3 含有 α^0 及 β^0 之旋转，也必含有 $\beta^0-h\alpha^0$ 之旋转．但 $0\leqslant\beta^0-h\alpha^0<\alpha^0$，而群不包含有小于 α^0 的正旋转．因此 $\beta^0-h\alpha^0=0$，即 $\beta^0=h\alpha^0$．特别是，由于群 K_3 包含 $360°$ 的旋转，则有一整数 n，使 $n\alpha^0=360°$，由之 $\alpha^0=\dfrac{360°}{n}$．

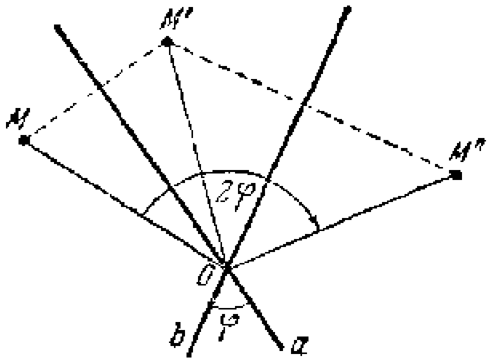


图 15

这样，群 K_3 由作 $0°$，$\dfrac{360°}{n}$，$2\,\dfrac{360°}{n}$，…，$(n-1)\dfrac{360°}{n}$ 的旋转组成．给 n 以值 2，3，4，…，我们得到群 K_3 的所有类型．在图 16 中给出的图形，它们的对称群是仅由所有绕 O 点作 $\dfrac{360°}{n}$，$n=19$，$n=3$ 的倍角的旋转组成．

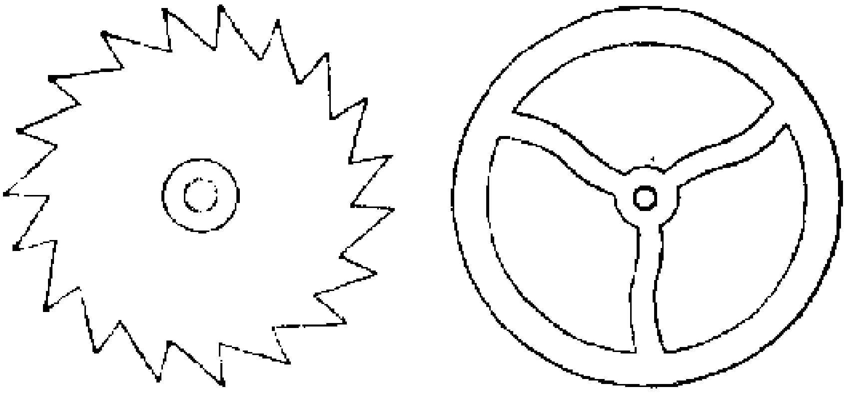


图 16

4. 只由一些旋转组成的对称群 K_4，但它含有可任意小的旋转. 此时转任意角 α 之旋转可由属于群 K_4 的旋转以任意精确度作出. 在这里我们有兴趣的当然仅是闭形，即含所有界点(参看第十七章 §9)的图形. 容易断定，对于闭形 K_4 包含作任意角的旋转. 这是有向圆对称的情形，附有转向的圆和环圈等等可作为它的例子(图 17). 这里在所有的容许变换下应当重合的不仅是图形本身还有转向，这就排除了关于直线的翻褶.

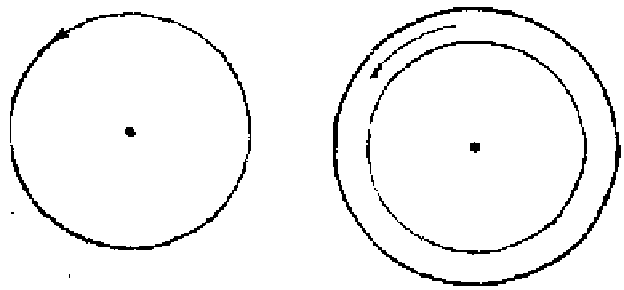


图 17

剩下来要考察的是混合的情形，当对称群 K 又含旋转又含翻褶. 我们略去证明，即使证明也是非常简单的，而只述结果：除去群 K_1—K_4 只存在下面两种类型.

5. 由 n 个关于过 O 点的直线系的翻褶，这些直线分平面为 $2n$ 个等角，及旋转组成的对称群 K_5，而这些旋转为作 $\frac{360^\circ}{n}$ 之倍角的旋转. 具有这样对称群的图形有，例如正 n 角形(图 18).

6. 由绕 O 点所有旋转及关于所有过 O 点的直线的翻褶组成的对称群 K_6. 这是完全圆对称的情形，无向圆及无向圆环可作为它的例子.

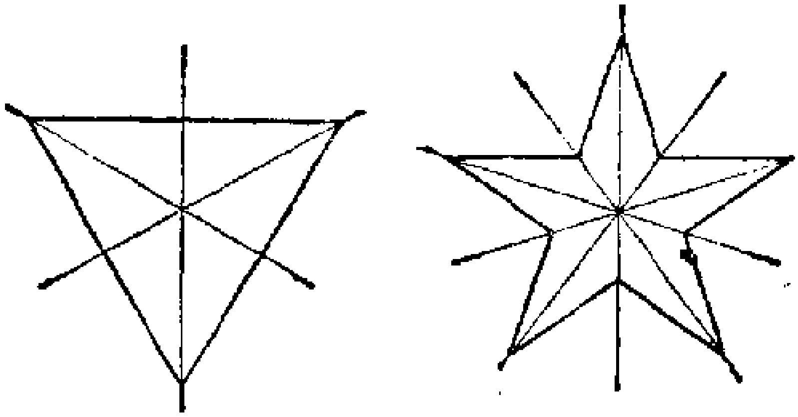


图 18

无限形的对称群 找出无限的平面图形的所有可能的对称群是较复杂的问题．当然在实际上永远不会给出全部无限平面．但是平面上的一块有时被非常小的图形盖满，和它们比较起来这一块就象是无限大了．例如，钢片的打磨得非常光滑的平面上却是被很小的、在显微镜下才看得清的花纹所布满．这些花纹反映了金属内部的结构的均匀性质．

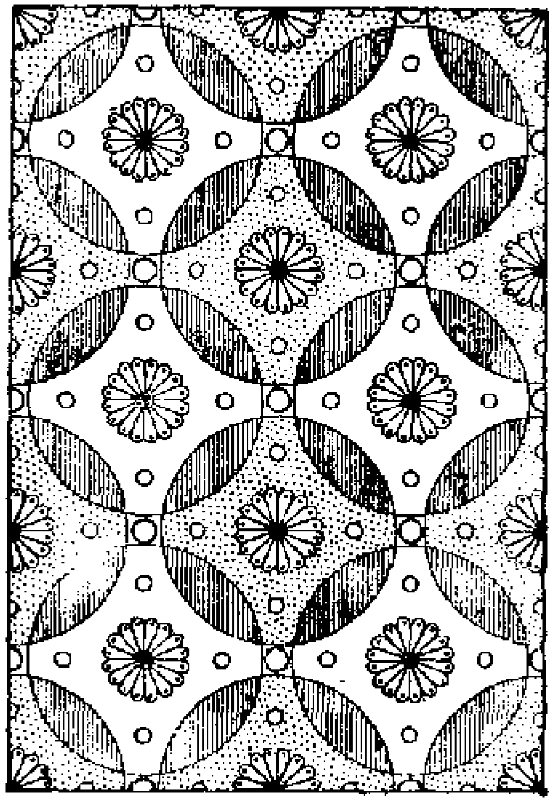


图 19

墙和布上的由重复图形作成的图案可作为另外一些例子．从古代开始一直到现在这种图案的艺术——装饰艺术——在很多民族广泛发展，图 19 中给出天花板上埃及图案的式样，它是纪元前一千五百年左右的作品．

就是对于有限形对称群我们已不得不分别考察情形 1, 2, 3, 5 及情形 4, 6, 在前者，对称群不含作任意小角的旋转，而在后者，群中有这样旋转．当研究无限形的对称群，特别是在空间的情形，这种把离散群和含有任意小变换的群分别讨论的作法有着更大的

意义．因此我们在一开始较确切地给出这两者之间的界限．

平面的运动群被称作是离散的，若对平面上每一点可圈以一个圆，使得群中每一运动或者使该点原位不动，或者把它一下子变出圆外．

和上面一样，可以找出所有离散的平面运动群．所有这些群都是平面图形的对称群，此处很自然地把离散的对称群分为三类：

I．在平面上有一点，它在所有变换下不动．这种类型的群包含上面举出过的群 K_1，K_2，K_3 和 K_4．

II．在平面上不存在不动点，但存在一直线，它在所有变换下仍和自己重合．此直线称为群的轴．具有这种类型的对称群的装饰品伸展成无限带（镶边）．这样的群一共有七个：

1．由一些平移作成的对称群 L_1，而这些平移所移动的距离为某一线段 a 的倍数．

2．由群 L_1 添加一个绕群轴上某一点作 180° 角的旋转所得的群 L_2．

3．由群 L_1 添加一个关于垂直群轴的直线的翻褶所得到的群 L_3．

4．由群 L_1 添加一个关于群轴的翻褶所得到的群 L_4．

5．由群 L_1 添加一个变换所得到的群 L_5，这个变换是关于轴的翻褶和作距离 $\frac{a}{2}$ 的平移的乘积．

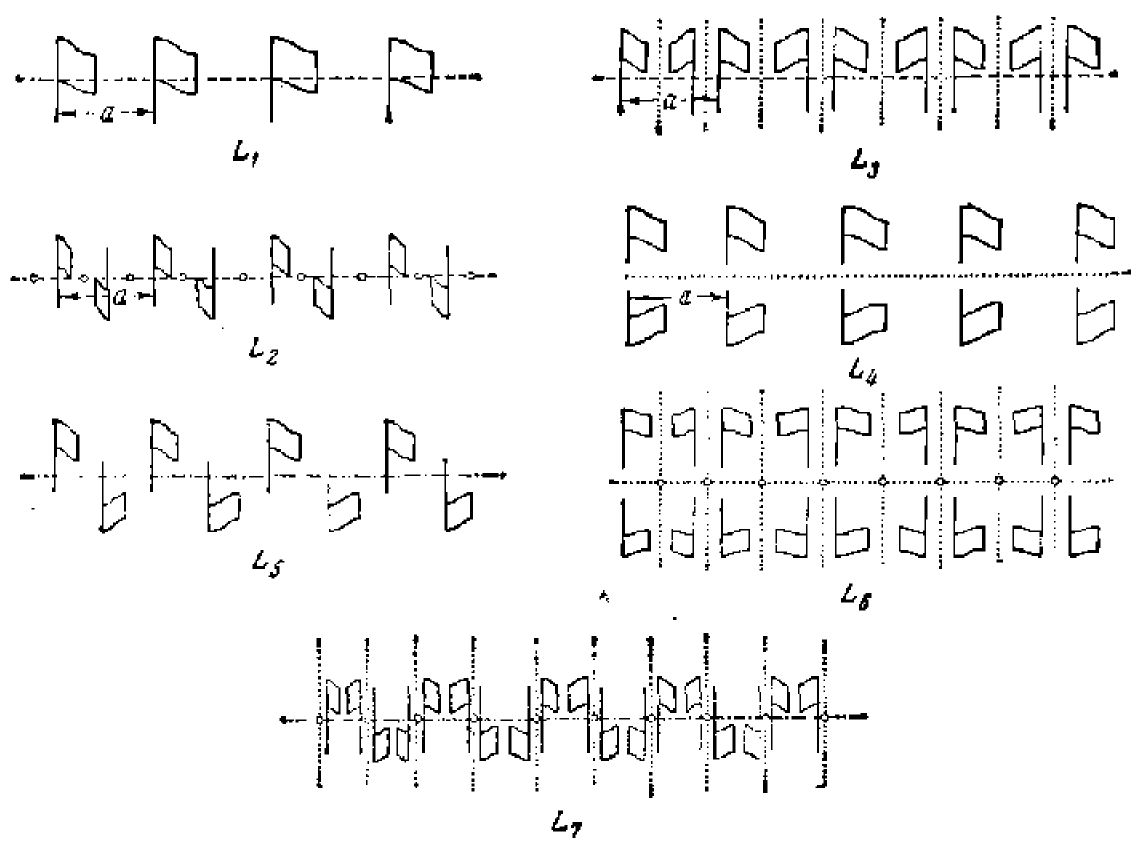
6．由群 L_4 添加一个关于某一垂直于群轴的直线的翻褶所得到的群 L_6．

7．由群 L_5 添加一个关于某一垂直于群轴的直线的翻褶所得到的群 L_7．

表 2 给出对应于每一群 L_1—L_7 的"镶边"的种种方案．

III．在平面上既不存在点也不存在直线，使它们在所有的变换下和自身重合．这种类型的群被称为平面费得洛夫群．它们是无限平面装饰品的对称群．它们共有 17 个：其中 5 个仅含第一种运动，12 个由第一种及第二种运动组成．

表 2



在表 3 中给出对应于 17 个平面费得洛夫群的装饰品图样，并且每一群由且仅由具下述性质之运动组成：它们把图中画出的小旗带到同一图中任意别的小旗上去。

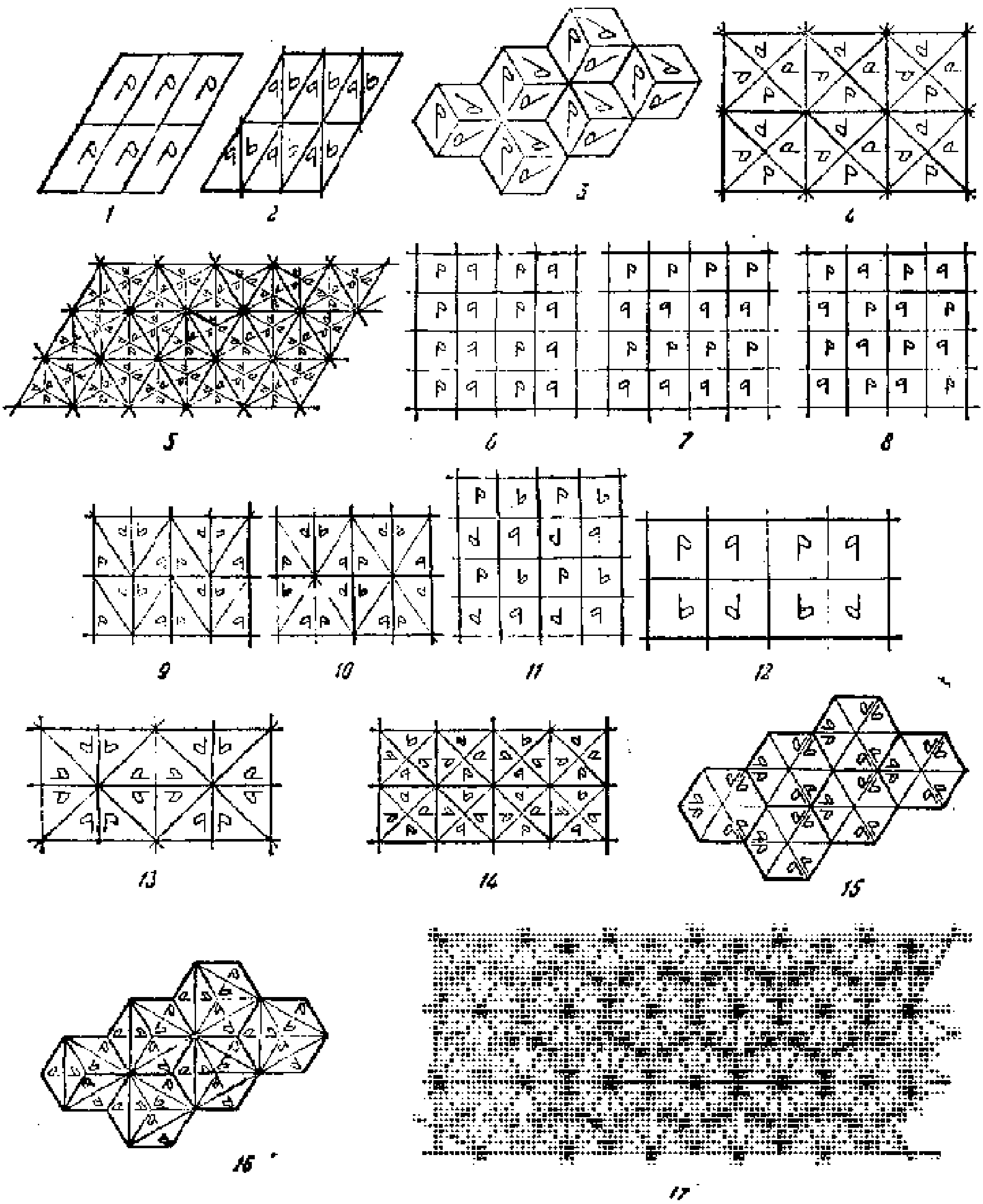
值得指出的是，装饰师实际上发现了具有所有可能对称群的装饰品，至于群论所作的只是证明不再有其他别的类型。

结晶群 在 1890 年著名的俄国结晶学家及几何学家费得洛夫用群论的方法解决了结晶学的基本任务之一，即规则的空间点系的分类问题。这是第一次直接应用群的理论去解决自然科学中重大问题，这对群论的发展起了重大的影响。

结晶体具有这样的特点，它们的组成原子在空间中组成某种意义下的规则体系。我们考察使体系中的点仍回到体系中点的空间运动。这些运动组成群，而该群的性质使得又能更精确的叙述点的规则体系这一概念。

空间的点系称作是**规则的空间点系**，若

表 9



1. 借助于使体系仍和自己重合的运动可把体系中任意一点带到它的任意其他别的点去.

2. 任一有限半径的球都不包含体系中无穷多个点.

3. 存在一正数 r, 凡是有半径 r 的任意球都必至少包含体系中的一个点.

研究结晶体的构造问题原来是和规则的空间点系分类紧密联系的，而后者又是和离散的空间运动群的分类相联系的．和平面情况相似，空间的运动群 H 被称作是离散的，若对每一点 A 可作一以 A 为心的半径为某一正数的球，使 H 中每一运动或使点 A 不动，或将它变到球外．

可以证明，使给定规则空间点系仍和本身重合的空间运动全体必是离散群，并且从体系的一个固定点出发，利用群中的变换来移动它可得到体系中所有的点．反过来，若已知某一离散群 H，则在空间中任取一点，利用 H 中所有的运动来推移此点，便可得一个点系，它具有性质 1，2，用加上不复杂的补充条件的方法，可以从离散群中划分出具有下述性质的群：它们对适当选定的点 A 给出真正的规则空间点系，即具有性质 1，2，3，的点系．这样的离散群叫做费得洛夫群或结晶群．从已说过的可看出，找出费得洛夫群是研究规则的空间点系的第一步和最重要的一步．原来，对于自然科学的目的而言，必须要考察的不仅是由固有运动组成的群，还有那些又含固有运动又含非固有运动的群(亦即，包含有翻褶)，仅由固有运动组成的费得洛夫群的个数大大地少于由固有运动及非固有运动组成的费得洛夫群的个数．仅在后面这个更一般的情形下，得到的规则空间点系的全体才真正地穷尽了在自然界中所遇到的晶体结构的全部丰富的内容．

值得指出的是，和上面提到的平面情形不同，只有群论才能帮助弄清这些非常众多的可能情况．

和平面情形比较，空间情形之复杂性可由下表看出：

空间费得洛夫群的数目

仅含第一种运动的群有 65 个

还含有第二种运动的群有 165 个

总共 230 个．

欲详细导出并列举所有的空间费得洛夫群，即使是在现在也还需要几十页的篇幅来叙述．因此，我们只限于介绍这些数量结

果，对此有兴趣的读者可参阅专门的文献[1].

结晶学的近代发展使得进一步发展对称的概念成为必要. 关于这方面的新的可能性和途径在结晶学家舒布尼柯夫院士的“有限体的对称性和反对称性”(苏联科学院出版, 1951) 一书中已指出.

§5. 伽罗华群

前几页所叙述的结果，为群论在解决晶体分类问题中所起的作用提供了一些概念. 然而这个问题并不是促使群论产生的原因，大约在这个问题产生的一百年以前，拉格朗日已经注意到代数方程根的对称性质与用根号解方程的可能性之间的联系. 在上世纪前三十年代著名数学家阿贝尔和伽罗华的工作中，对这种联系作了深入的研究，致使关于代数方程用根号可解性条件的著名问题得到解决. 这一问题的解决完全依赖于对置换群的性质作精细的讨论，而这就是群论的实际起源.

研究代数方程的性质与群的性质之间的联系，已成为现代一门巨大理论——伽罗华理论的对象.

关于这一问题的历史和伽罗华理论的意义，已在第四章(第一卷)中叙述过了. 不过，因为伽罗华理论在群论的发展中起着决定性的作用，所以我们这里还要提一提这一理论的基本事实，不过在提法上现在采用了最适宜于解释群论的形式. 这些事实的证明，需要用到许多辅助概念，因此把它们略去.

代数方程的群　设给了一个 n 次方程

$$x^n+a_1x^{n-1}+\cdots+a_n=0, \tag{6}$$

1) 在 D. 希尔伯特和 C. Э. 阔恩-佛新所著“直观几何学”(莫斯科-列宁格勒, 1951)一书中，详细地叙述了如何导出由第一种运动组成的平面离散群. 关于空间结晶群的导出，可在 E. C. 费得洛夫的重要著作“规则体系的对称”(费得洛夫，基本论文，苏联科学院出版，1949) 中及 B. 杰龙涅，H. 巴都洛夫，A. 亚历山大洛夫所著“结晶结构分析的数学基础”(莫斯科-列宁格勒，1934)，或 C. A. 波哥莫洛夫所著“按费得洛夫方法规则体系的导出”(库布奇出版，1934) 等书中找到.

它的系数看作是给定的数值，例如是一些复数. 由方程系数的有限次加、减、乘、除运算可能得到的一切数的集合，称为方程的**基本域**或方程的有理域.

例如，若方程有有理系数，那么它的有理域是由全体有理数组成的；若方程是 $x^2+\sqrt{2}x+1=0$，那么它的有理域由一切下述形状的数组成：$a+b\sqrt{2}$，其中 a, b 是有理数.

今以 $\xi_1, \cdots, \xi_n$ 表示给定方程的根. 由这些根实行有限次加、减、乘、除运算可以得到的一切数的集合，称为方程的**分解域**. 例如，方程 $x^2+1=0$ 的分解域是有有理系数的复数 $a+bi$ 组成的集合，而上面提到的方程 $x^2+\sqrt{2}x+1=0$ 的分解域，则是由下述形式的数组成的集合：$a+bi+c\sqrt{2}+di\sqrt{2}$，其中 a, b, c, d 是有理数.

根据维达公式，方程的系数可以由它的根实行加、乘运算得到，所以方程的分解域永远包含它的基本域. 有时这两个域是重合的.

分解域到它自身上的一个一一映象 A，叫作分解域关于基本域的一个自同构，假如对于分解域的每一对元素，它们的和映射到和，积映射到积，并且基本域的每一元素映射到它自身. 上面所说的性质可以写成公式

$$(a+b)A=aA+bA,\ (ab)A=aA\cdot bA,\ \alpha A=\alpha$$
$$(a,\ b\in K\quad \alpha\in P), \tag{7}$$

这里 aA 是在映象 A 下元素 a 所映射到的元素，P 是基本域，K 是分解域.

按照 258 页所说的一般原则，分解域关于基本域的所有自同构的集合作成一个群. 这个群叫作给定方程的**伽罗华群**.

为了给出关于伽罗华群比较具体的表示法，首先要注意，伽罗华群中的自同构将给定方程的根仍然映射到这个方程的根. 事实上，如果 x 是方程(6)的一个根，那末对方程的两端实行自同构 A，并且应用性质(7)，我们得到

$$(xA)^n+a_1A(xA)^{n-1}+\cdots+a_nA=0\cdot A,$$

因为 $0\cdot A=0$, $a_iA=a_i$, 所以有

$$(xA)^n+a_1(xA)^{n-1}+\cdots+a_n=0,$$

而这就是所要证明的. 因此, 每一个自同构都导出方程根的集合的一个确定的置换. 另一方面, 知道了一个这样的置换, 由于分解域的所有元素都可由根经过算术运算得出来, 因此我们也就同时知道了一个自同构. 这证明了, 为了讨论自同构群, 可以讨论与它相当的方程根的置换群. 由此可以推出, 所有伽罗华群都是有限群.

求任意方程的伽罗华群通常是一个复杂的问题, 只有在个别情况下, 伽罗华群才是比较容易求出的. 例如系数为 $a_1, \cdots, a_n$ 的方程 (6), 这个方程的有理域是由系数的有理分式(即分子和分母是 $a_1, \cdots, a_n$ 的多项式的分式)组成的. 分解域是由方程的根 $\xi_1, \cdots, \xi_n$ 的有理分式组成的, 这里根与系数的关系有公式

$$\begin{aligned}-a_1&=\xi_1+\xi_2+\cdots+\xi_n,\\ a_2&=\xi_1\xi_2+\xi_1\xi_3+\cdots+\xi_{n-1}\xi_n,\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ (-1)^na_n&=\xi_1\xi_2\cdots\xi_n.\end{aligned}\tag{8}$$

因为方程 (6) 是"一般的", 所以我们可以把它的根看作独立的变量. 于是这些根的所有置换都将导出分解域的自同构. 公式 (8) 表明, 在任一这样的自同构之下, 系数映射到自身, 因而同时它们的有理分式也就映射到自身. 这样一来, 一般 n 次方程的伽罗华群是 n 个文字的所有置换的对称群.

也可以举出这样数值系数的方程, 它的伽罗华群是对称群. 例如可以证明, 对于任意的 n 来说, 方程

$$1-\frac{n}{1}x+\frac{n(n-1)}{1\cdot2}\cdot\frac{1}{1\cdot2}x^2-\frac{n(n-1)(n-2)}{1\cdot2\cdot3}\cdot\frac{1}{1\cdot2\cdot3}x^3+\cdots+(-1)^n\frac{1}{1\cdot2\cdots n}x^n=0\tag{9}$$

的伽罗华群就是 n 次对称群.

一般已经知道造一个方程使它以预先给定的群作为伽罗华群的方法. 不过这需要在这样的条件下, 方程的系数可以任意选取.

如果要求建立的方程有有理系数的话，那么现在只知道对于个别类型的群的造法．苏联数学家涉法列维奇在这方面获得了值得称道的成就，他得到建立有理系数方程以预先给定的可解群为伽罗华群的方法．但是对于一般情况，这一问题仍未能解决．

方程用根号的可解性 由定义看到，方程的伽罗华群刻划出方程根的内在对称性．可以证明，所有最重要的问题，如判别把解给定方程化为解一个较低次方程的可能性以及许多其他问题，都可以用伽罗华群的结构问题的形式叙述出来，而每一个 n 次方程的伽罗华群都是某一 n 次置换群，即总共有有限个对象．至少在理论上，它们的所有关系都可以用试验的方法找出来．

研究伽罗华群是解决与高次代数方程有关问题的一种值得重视的方法．例如可以证明，当且仅当伽罗华群是可解群时，方程可用根号解（可解群的定义见 § 3，272 页）．以前曾经指出，2，3，4 次的对称群是可解群．这就找到了 2，3，4 次方程可用根号解这个一般事实的全部根据．5，6 等次的“一般”方程的伽罗华群是对称群，其次数与方程的次数相同，这些群是不可解群．所以推出，4 次以上的一般方程不能用根号解．

至于不能用根号解的数值方程可取方程(9)($n>4$)为例，因为它的伽罗华群也是对称群．

§ 6. 一般群论的基本概念

在十九世纪，群论主要发展了变换群的理论．但是渐渐地完全明白了，那里所得到的最重要的结果只是依赖于变换可以相乘，及依赖于这个运算具有一系列的特性．另一方面，可以找到一些事物，它们并不是变换，但是对它们却可以实行某种运算(不妨称之为乘法)，这个运算具有同变换群的乘法一样的性质，并且可以证明，变换群的主要定理对它们来说也都成立．因此近一百年来群的概念已有力地运用到任意元素的系统上面，而不仅是变换的系统．

群的一般定义 目前一般采用的群的定义如下：设任一集合 G 的每一对按一定顺序给出的元素 a, b, 都规定了这个集合中一个确定的元素 c 与它们对应. 这时我们说，在集合 G 上给定了一个运算. 通常对运算我们采用一个一定的名称，如加法、乘法、结合法等. G 中与元素 a, b 对应的元素，在这种情况下，分别称为元素 a, b 的和；a, b 的积；a, b 结合的结果；并分别记作 $a+b$, ab, $a*b$. "加法"或"乘法"这种名称可以用于与数的普通加法和乘法毫不相干的情况.

集合 G, 连同定义在它上面的运算 $*$, 叫作对于这个运算来说的一个群，假如满足下面群的公理：

1. 对于 G 的任意三个元素 a, b, c 来说，有

$$x*(y*z)=(x*y)*z \text{（结合律）}.$$

2. G 的元素中有一个元素 e, 对于 G 的任何元素 x 来说，有

$$x*e=e*x=x.$$

3. 对于 G 的任一元素 a 来说，在 G 中存在一个元素 a^{-1}, 使得

$$a*a^{-1}=a^{-1}*a=e.$$

公理 2 提到的元素 e, 称为群的恒等元. 公理 3 所断言存在的元素 a^{-1}, 称为 a 的逆元. 如果群的运算叫作加法或乘法时，那么恒等元就相应地叫作零元或单位元，而群的公理就有下面的形式：

1) $x+(y+z)=(x+y)+z$, 1) $x(yz)=(xy)z$,

2) $x+0=0+x=x$, 2) $xe=ex=x$,

3) $x+(-x)=(-x)+x=0$, 3) $xx^{-1}=x^{-1}x=e$.

在前几节中已看到许多群的例子. 这些群的元素是变换，群的运算是变换的乘法. 数 0, ±1, ±2, …的集合对于加法运算来说也是一个群，因为整数的和仍是一个整数，整数的加法是结合的，整数零是恒等元，并且对于所讨论的数中的每一个数 a, $-a$ 是逆元. 另一个群的例子是由所有实数（零除外）的集合对于乘法所构成的. 事实上，两个非零实数的积仍是一个非零实数，实数的

乘法运算是结合的，有恒等元等于1，每一非零实数a有倒数$a^{-1}=\frac{1}{a}$. 可以不断地举出类似的例子.

虽然群的运算可以有不同的名称，不过我们约定以后几乎永远把它叫作乘法. 子群、群的元素的幂，循回群，群的元素的周期等概念，可以像对变换群一样来定义(参看§3)，我们不再重复. 这里要指出的只是，G的元素a称为与G的元素b共轭，假如在G中存在一个元素x，使得$b=x^{-1}ax$. 因为$a=a^{-1}aa$，所以群的每一元素都与它自己共轭. 又由$b=x^{-1}ax$，显然推出$xbx^{-1}=a$或$a=(x^{-1})^{-1}bx^{-1}$，这就是说，如果元素a与b共轭，则b也与a共轭. 最后，若$b=x^{-1}ax$且$c=y^{-1}by$，则

$$c=y^{-1}x^{-1}axy=(xy)^{-1}a(xy).$$

所以，若两个元素分别与同一第三元素共轭，那么这两个元素共轭. 这些性质证明了，在一个群中，所有的元素可以分成由相互共轭的元素组成的类. 不过，若是群可换，即对于任何x及y有$xy=yx$，那么共轭的元素是相等的，因此上述的每一共轭类，都含且只含一个元素.

同构 在群的概念中可以分为两个方面. 为了给出一个群，必需：1) 指出它的元素是那些事物，2) 指出元素相乘的规则. 依此，研究群的性质可以在两个不同的观点下进行. 可以研究群的元素和元素的集合的各种性质同它们关于群的运算的性质之间的联系. 这种观点常常在研究各个具体的群，例如，在研究空间或平面的运动群的时候被采纳. 另外，也可以研究群的这样的性质，这些性质可以完全由群的运算的性质表示出来. 这种观点说明了抽象群论或一般群论的特性，它可以用同构的概念说明得更清楚.

称两个群是同构的，假如其中一个群的元素可以同另一群的元素对应起来，使得第一个群中任何元素的积对应到第二个群中相应元素的积. 两个群之间具有上述性质的一一对应，称为一个同构对应.

容易看出，两个群的元素，若在同构对应之下相互对应着，则

对于群的运算来说具有相同的性质. 例如,在同构对应下,一个群的恒等元、互逆的元素、具有给定周期 n 的元素、子群,分别对应到另一个群中的恒等元、互逆的元素、具有同一周期 n 的元素、子群. 所以, 可以说抽象群论是只研究群的那些在同构对应下保持不变的性质的. 例如,按照抽象群论的观点,四个元素的所有置换作成的群同空间把一个固定的正四面体变为自身的固有运动和非固有运动作成的群具有相同的性质,因为它们是同构的. 事实上,上述的运动把四面体的四个顶点仍然变为它的顶点. 这样的运动的个数等于 24. 对每一个运动,令由它导出的顶点的置换与之对应,我们得到这两个群之间的一一对应,而这就是所求的同构对应.

对数的理论给出一个同构对应的有意义的例子. 对于每一正实数,令这个数的对数与之对应,我们得到正实数集与所有正负实数的集合间的一个一一对应. 关系式 $\log(xy)=\log x+\log y$ 证明这样建立的对应是正实数集对乘法作成的群到所有实数对加法作成的群上的一个同构对应. 这个同构对应的实际重要性是众所周知的.

元数不同的有限群可以作为不同构的群的例子.

按照上面所说的,抽象群由它的元素的乘法规则定义,所以与它的元素的个性无关. 因此,同构的各个群中的一个具体给定的群可以当作这些抽象群的一个模型.

一个抽象群可以用各种方式给出,其中最自然的,至少对有限群来说,是用"乘法表"来给出.

对于元数为 n 的群来说, 把它的元素按照任意的顺序写好,那么它的乘法表是由分成 n 行 n 列的一些方格构成的. 位于第 i 行第 j 列的格子里的元素是第 i 个元素与第 j 个元素的积. 有限群的乘法表也有时叫作它的凯利表.

但是由于过分繁琐, 利用乘法表来实际地给出群几乎是不被采用的.

还有其他的方法来给出抽象群. 我们常用的一种是,利用生成元素组及定义关系式来给出群. 但是抽象群常常是由给出一个与

它同构的具体群，特别是变换群，确定出来.

自然会发生这样的问题：是否任一抽象群都可以看作一个变换群. 下面的定理回答了这个问题，每一个群 G 都与它的元素集合的某一个变换群同构.

事实上，设 g 是 G 的一个固定元素，用 A_g 代表集合 G 的这样一个变换：在这个变换之下，G 的每一个元素 x 有元素 xg 与之对应. 由于对于每一给定的元素 a，方程

$$xA_g = xg = a$$

有唯一解 ag^{-1}，所以变换 A_g 是一一的. 另一方面，群的元素的乘积 gh 有变换的乘积 A_gA_h 与之对应，因为

$$xA_{gh} = x(gh) = (xg)h = (xA_g)A_h = x(A_gA_h).$$

群 G 的恒等元 e 有恒等变换与之对应，而逆元 g^{-1} 有逆变换与之对应. 所以，对应于 G 的元素的所有变换的集合 Γ 是一个与 G 同构的变换群. 容易相信，如果群 G 的元素大于 2，集合 Γ 不包含集合 G 的所有变换，所以 Γ 只是集合 G 的所有变换的“对称群”的一个子群.

不变子群和商群 设 P 和 Q 是任意两个由某一群 G 的元素组成的集合. 群 G 中一切可以表为 P 的某一元素与 Q 的某一元素乘积的元素所成的集合，叫作集合 P 与集合 Q 的乘积，记作 PQ. 特别，乘积 gP（其中 g 是群 G 的一个元素）是 g 与集合 P 的每一元素的乘积组成的集合.

如果群 G 的一个子群 H，对于 G 的每一元素 g 来说，有 $gH = Hg$，则称 H 为 G 的一个**不变子群**或**正规子群**. 形如 gH 和 Hg 的集合（其中 H 是任一子群），分别称为群 G 对子群 H 含元素 g 的**左陪集**和**右陪集**. 这样一来，我们可以说，不变子群完全由下述性质描述出来：对于不变子群来说，含同一元素的左陪集和右陪集是重合的.

如果 H 是一个不变子群，那么容易证明，两个陪集的乘积仍是一个陪集，且 $aH \cdot bH = ab \cdot H$. 子群 H 本身也是一个陪集，即对应于恒等元 e 或对应于它的任一元 h 的一个陪集，因为 $hH = H$.

陪集的乘法是结合的：

$$(aH\cdot bH)\cdot cH=(ab\cdot c)H=(a\cdot bc)H=aH(bH\cdot cH).$$

在这一乘法中，子集 H 起着恒等元的作用：$H\cdot aH=eH\cdot aH=(ea)H=aH$. 类似地也有 $aH\cdot H=aH$. 陪集 $a^{-1}H$ 是 aH 的逆元，因为 $aH\cdot a^{-1}H=a^{-1}H\cdot aH=H$. 因此，用对于某一不变子群来说的所有陪集作元素组合成一新的集合，这个集合对陪集的乘法来说是一个群. 这个群叫作群 G 对不变子群 H 来说的商群，记作 G/H.

容易断定，有限群对于任一子群 H 来说的每一个陪集中含有元素的个数都与子群 H 含有元素的个数相等，且不同的陪集不含公共元素. 所以推出，有限群 G 对它们的任一子群 H 来说的陪集的个数，等于 G 的元素个数被 H 的元素个数除所得的商. 由此得出重要的拉格朗日定理：有限群的任一子群的元数是群的元素的一个因数.

由不变子群的定义易见，阿贝尔群的所有子群都是不变子群. 另一种极端的情形是所谓的单群，它除了单位子群和本身以外，没有任何一个其他的子群是不变子群. 除阿贝尔群和单群外，可解群也有重要的意义，可解群的定义已在 § 3 给出. 可以证明，可解群有一个不变子群的有限链 $G, G_1, G_2, \cdots, G_k$，其中第一个就是给定的群，每一个后面的子群都被包含在前一个子群里面，最后的子群是单位群，并且所有的商群 $G/G_1, G_1/G_2, \cdots, G_{k-1}/G_k$ 都是阿贝尔群.

同态商群的概念与全部群论基础的同态映象概念有着非常密切的联系. 群 G 元素的一个集合到群 H 元素的一个集合上的单值映象叫作一个同态对应或同态映象，如果第一个群中每两个元素的乘积映射到第二个群中相应元素的乘积.

这样一来，对于群 G 的元素 x 如果用 x' 表示群 H 中与之相应的元素，则同态映象可以用下面的性质描述出来：

$$(x_1x_2)'=x_1'x_2'.$$

由同态对应和同构对应的定义，显然知道，同构对应必须是一

一的，而同态对应则只有一方面是单值的，即对群 G 的每一元素在 H 中有唯一元素与之对应，但是 G 的不同元素在 H 中可以有同一对象．按通常的意义可以断言，在同构对应之下，群 H 是群 G 的一个精确的写照，而在同态映象之下，当把 G 映射到 H 时，G 的某些元素的差异被抹掉了，而某些元素"聚合"到 H 的一个元素上面．然而同态映象的这种"粗糙性"并不是一个缺陷，相反地倒是一个优点，致使同态映象能够用来作为研究群的性质的一个有力工具．

在许多情况下，同态映象常伴随变换自然而然地呈现出来．例如正四面体的对称群（图 20）．这个群与四个元素置换的对称群同构，因为存在一个唯一的运动（第一种的或第二种的），把顶点 A_1, A_2, A_3, A_4 变成另一个任意给定的排列．

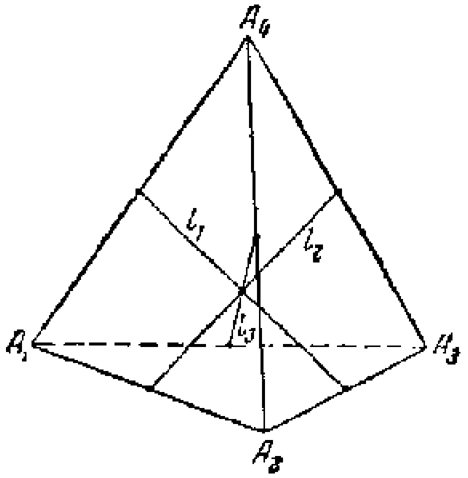


图 20

现在看联结对棱中点的直线 l_1, l_2, l_3．每一个把四面体变成自身的运动，都产生一个 l_1, l_2, l_3 的置换，并且 l_1, l_2, l_3 的任一置换也都产生一个四面体的对称．显然，四面体的变换的乘积对应直线 l_1, l_2, l_3 的置换的乘积．由图 20 容易看出，建立四元素 A_1, A_2, A_3, A_4 的置换的对称群到三元素 l_1, l_2, l_3 的置换的对称群上的一个同态映象的自然方式．也不难找出"大群"中在这个同态映象下"聚合"到一起的元素．

再看一些一般的例子．n 个元素的所有置换的集合，当 $n>2$ 时构成一个不可换群．另一方面，数 $+1$ 和 -1 对数的乘法来说也构成一个群．现在令任意 n 个元素的每一个偶置换对应到 $+1$，而每一奇置换对应到 -1．这样就给出了一个 n 元素置换的对称群到群 $\{+1,\ -1\}$ 上的同态映象，因为根据 § 3，奇偶性相同的置换的乘积是一个偶置换，而奇偶性不同的置换的乘积是一个奇置换．

另一个例子：如果每一实数 $x\neq 0$ 用它的绝对值 $|x|$ 与之对应，那么所产生的正负实数(零除外)对乘法来说的群到只由正实数对乘法作成的群上面的映象是一个同态对应，因为 $|xy|=|x||y|$.

以前曾经提到，平面上每一个第一种运动 A 可以表成一个适当的绕固定点 O 的旋转 V_A 与某一平移 D_A 的乘积. 绕定点 O 的旋转作成一个群. 所以对应 $A\to V_A$ 把平面上第一种运动的群单值地映射到绕点 O 的旋转群上. 今证这个映象是一个同态映象. 由分解式 $A=V_AD_A$; $B=V_BD_B$ 推出

$$AB=V_AD_AV_BD_B=(V_AV_B)(V_B^{-1}D_AV_BD_B).$$

第一括号中是绕 O 点的一个旋转，而第二括号中是平移变换 $V_B^{-1}D_AV_B$ 与平移变换 D_B 的乘积，因而是一个平移. 这就证明了运动的乘积 AB 对应到相应旋转的乘积 V_AV_B，即所说的映象是一个同态映象.

最后证明，任一群对它任意一个正规子群 N 来说的商群 G/N 是群 G 的一个同态象.

事实上，对群 G 的每一元素 g 令含 g 的陪集 gN 与之对应，我们就得到一个所求的 G 到 G/N 上的同态映象，因为乘积 gh 有陪集 ghN 与之对应，而 ghN 等于与元素 g 和 h 对应的陪集 gN 和 hN 的乘积.

回到同态映象的一般性质，可以证明，在任意同态映象之下，恒等元映射到恒等元，并且互逆元素映射到互逆元素.

事实上，如果 e 是群 G 的恒等元，e' 是它在 H 中的象，则由 $ee=e$ 推出 $e'e'=e'$，用 ε 表示群 H 的恒等元，得 $e'=e'e'^{-1}=\varepsilon$. 第一论断证完. 现在令 x 和 y 在 G 中是互逆元素，而 x' 和 y' 是它们在 H 中的象. 由 $xy=e$ 推出 $x'y'=e'=\varepsilon$，即 x' 和 y' 在 H 中是互逆元素，并且可知

$$(x^{-1})'=x'^{-1}.$$

所证明的论断，保证我们容易找出群 G 中任意元素乘积的象. 例如

$$(ab^{-1}c^{-1}dh^{-1})'=a'(b^{-1})'(c^{-1})'d'(h^{-1})'=a'b'^{-1}c'^{-1}d'h'^{-1}.$$

下面的定理在全部同态映象的理论中居于基本地位。

在任意群 G 到群 H 的一个同态对应下，群 G 中所有对应到 H 中恒等元 e' 的元素集合 N，是 G 的一个不变子群，而对应到群 H 任意一个固定元素的 G 中元素集合，是 G 对 N 的一个陪集，并且这样建立起来的 G 对 N 的陪集与群 H 的元素间的一一对应，是 H 与商群 G/N 间的一个同构对应.

现在证明这个定理. 令 a, b 是 N 的任意两个元素，这表示 $a' = b' = e'$，这里以及此后，加撇表示 G 的元素在 H 中的象. 但

$$(ab)' = a'b' = e'e' = e',$$

$$(a^{-1})' = a'^{-1} = e'^{-1} = e',$$

即 ab 和逆元 a^{-1}, b^{-1} 属于 N，所以集合 N 是一个群. 其次，对于 G 的任意元素 g,

$$(g^{-1}ag)' = g'^{-1}a'g' = g'^{-1}e'g' = g'^{-1}g' = e',$$

即对 G 的任意元 g 和 N 的任意元 a 来说，$g^{-1}ag$ 属于 N. 由此显然推出，N 是 G 的一个不变子群. 定理的第一个论断证完.

为了证明定理的第二个论断，在 G 中任取一元 g，并且考虑 G 的所有这样元素 u 的集合 U，这些元素的象 u' 与元素 g 的象 g' 重合. 设 $u \in gN$，即 $u = gn$，其中 $n \in N$，于是 $u' = g'n' = g'e' = g'$. 所以 $gN \subset U$. 反之，若 $u' = g'$，则 $(g^{-1}u)' = g'^{-1}u' = g'^{-1}g' = e'$，即 $g^{-1}u = n$，这里 n 是 N 的一个元素. 由此 $u = gn$，故知 $U \subset gN$. 由 $gN \subset U$ 及 $U \subset gN$ 得出 $U = gN$.

最后，定理的第三个论断是显然的：商群 G/N 中的任意两个陪集 gN, hN 对应到 H 的元素 g', h'，而由公式

$$gN \cdot hN = ghN,$$

陪集的乘积对应到 $(gh)' = g'h'$，这就是所要证明的。

这个关于同态对应的定理证明了群 G 的每一个同态象 H 与相应的商群 G/H 同构. 这样一来，在同构的观点下，给定的群 G 的所有同态象被它的各个商群包罗净尽.

§7. 连续群

李群．连续变换群 在解高次代数方程时，群论所获得的成就，促使上世纪中叶的数学家们企图把群论用来解其他类型的方程，首先是用来解在数学应用中起着非常巨大作用的微分方程．这个企图终于实现了．尽管群在微分方程里，较之在代数方程论里，处于完全不同的地位，但应用群论来解微分方程的研究，却使得群的观念本身有了本质上的扩充，而建立了所谓连续群或李群的新理论．连续群对于形形色色的数学部门的发展说来，也是极为重要的．

可是代数方程的群只由有限多个变换组成，而以类似方法所构造的微分方程的群却是无限的．此外，属于微分方程群的变换，可以借助于有限参数组来给出，改变这些参数的数值，可以得到群的所有变换．例如，假设群的所有变换是由参数 $a_1, a_2, \cdots, a_r$ 来界定．使这些参数具有值 $x_1, x_2, \cdots, x_r$ 时，我们就得到一个变换 X；使同一些参数具有新值 $y_1, y_2, \cdots, y_r$ 时，得到另一个变换 Y．按照这些变换相乘的条件，$Z=XY$ 是群的一个变换，这就是说，我们得到了确定的新参数值 $z_1, z_2, \cdots, z_r$．值 z_i 依赖于 $x_1, x_2, \cdots, x_r; y_1, y_2, \cdots, y_r$，即得到了依赖于这些参数值的一些函数：

$$z_1=\varphi_1(x_1, x_2, \cdots, x_r; y_1, y_2, \cdots, y_r),$$
$$z_2=\varphi_2(x_1, x_2, \cdots, x_r; y_1, y_2, \cdots, y_r),$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$z_r=\varphi_r(x_1, x_2, \cdots, x_r; y_1, y_2, \cdots, y_r).$$

如果一个群的元素连续依赖于有限参数组的值，而乘法律是借助于可微分两次的函数 $\varphi_1, \varphi_2, \cdots, \varphi_r$ 来表示，那么这个群就叫做李群．这是以首先研究这种群的挪威数学家索福斯·李来命名的．

十九世纪前半期，罗巴切夫斯基讨论了一种新的几何系统，这种几何学现在就叫做罗巴切夫斯基几何．大约在同时投影几何形成了一个独立的几何系统；稍后建立了黎曼几何．因此，到十九世

纪后半期，从研究"现实世界的空间形式"(恩格斯)的种种观点出发，可以产生一系列独立的几何系统．要把所有这些系统在同一个观点下统一起来，但要保存它们最重要的质的差异，是可以借助于群论来实现的．

考虑某个几何空间点集的一一变换，它们不改变已知几何系统中图形间的一些基本关系．这些变换的全体组成一个群，通常叫做已知几何系统的运动群或自同构群．运动群完全刻划了已知几何系统，因为若运动群已知，那么相应的几何学就可以看成是一种学问，它研究点集在给定群的变换下保持不变的性质．不同的几何系统按其运动群的分类方法，是在上世纪后半期由克莱茵提出的．关于这种方法以及各式各样的几何系统，在抽象空间一章里已经有所叙述．这里我们只想提一下，所有在上世纪实际研究过的几何系统，它的运动群都是李群．由于这个缘故，研究李群的课题就显得特别重要．

由于和数学与力学的许多不同分支的广泛联系，使得李群理论从它的建立到如今，一直是在飞跃发展着．并且发生这种现象，对于有限群说来现在也还没有解决，但对李群却较为迅速地得到解决，即单纯有限群(即不具有非当然不变子群的有限群)的分类问题，直到现在仍然少有进展，而单纯李群相应的分类问题，还在上世纪末叶就由基林和卡尔丹解决了．在发展李群理论中，苏联数学家莫洛佐夫、马尔采夫、邓肯完全解决了李群的单纯子群分类问题，这是长久期待解决的一个重要问题．李群理论在另一方向被苏联数学家盖尔芳特和奈玛尔克发展起来了．他们发现了单纯李群可以用希尔伯特空间的酉变换来既约表示；后一问题对于分析和物理特别有意义．

李群的研究是借助于一种独特的工具，所谓"无穷小群"或李代数来进行的．详见§13．

拓扑群　在苏联除了李群的古典理论的巨大发展外，更一般的拓扑群或连续群理论获得了突出的成就．李群是要求群的元素以有限参数组来界定，并且乘法律是借助于可微函数来表示．和

李群概念不同，拓扑群概念更简单而广泛。即是说，一个群叫做**拓扑群**，如果对于它们的元素说来，除了通常的群运算外，还定义了邻近概念，使得从群的一些元素邻近，就推出它们的乘积的邻近，它们的逆元素也邻近。

最初，为了把李群理论的很多基本概念整理得有条理，拓扑群的观念是必需的。但是后来就显露出来，这个概念对于另外一些数学部门也极为重要。关于一般拓扑群论的初步工作，是在本世纪 20 年代开始，但能使人们承认是出现了一个新的分支的那些基本结果，还是在 20 年代末到 30 年代初才被得到。相当大一部分结果是被苏联数学家邦特里雅金获得的，因此他不愧被认为是近代拓扑群论的奠基者之一。邦特里雅金的著作"连续群"包含着在世界文献中关于连续群理论的第一个总结性论述，在这个领域里已经将近 20 年仍旧是基本的文献。

§8. 基 本 群

在前面几节的所有已经考虑过的具体情况下，群通常是作为这种或那种集合的变换群。仅有对于加法或乘法所成的数群是例外。现在我们来分析一个重要的例子，这个群的最初发现不是变换群而是作为带有一个运算的代数系统。

基本群 我们考虑某一个曲面 S 和在它上面的一个动点 M。使 M 在曲面上跑过一条连结点 A 和点 B 的连续曲线，我们就得到一条由 A 到 B 的确定的道路。这条道路可以自己相交任意次数甚至可以在一些个别的地段里来回跑。为了指明一条道路，只给出点 M 在它上面移动的曲线是不够的。还需要指明，这个点以某方向和某些次数所重复跑过的那些地段。例如，点可以以不同的次数和不同的方向跑过同一圆周，而且所有这些循环道路都认为是不同的。两条带有同一个起点和终点的道路称为是等价的，如果两条道路中的一条可以经一连续的形变变成另外一条。在平面或者球面上的任意两条连结点 A 和点 B 的道路都是等价的

(图 21)．然而在环面上，例如，由点 A 出发又回到点 A 的两条闭道路 U 和 V(图 22)彼此不等价．如果代替环面而考虑在两端是无限伸展的圆柱面并且在它上面取一条道路 X(图 23)，那么容易理解，在圆柱面上的任意带有起点 A 的闭道路将等价于型为 $X^n(n=0,\ \pm1,\ \pm2,\ \cdots)$的道路，这里的 $X^n(n>0)$应了解为道路 X 重复了 n 次而得的道路；X^0 了解为仅由一个点 A 组成的零道路，而 X^{-n} 了解为以相反的方向跑过 X^n，例如，$Z\sim X^{-1}$，$Y\sim X^2$，$U\sim X^0$(图 23)．这个例子指明了道路等价概念的意义：在圆柱面上的不同的闭道路组成一无法描述的集合，但在等价的意义下所有这些道路都归结为以这样的或那样的方向和以足够的次数跑过圆周 X．当 $m\neq n$，道路 X^m 和 X^n 不等价．

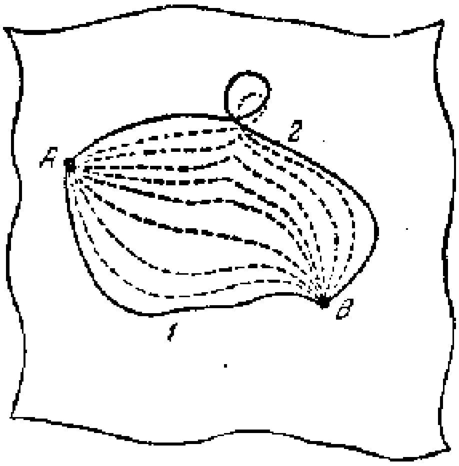


图 21

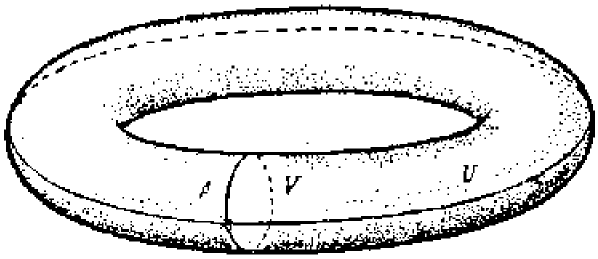


图 22

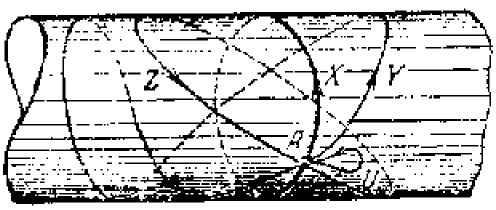


图 23

现在转回来考虑任意的曲面．我们假设在它上面给了两条道路 U 和 V，道路 U 是由点 A 到点 B，而道路 V 是由点 B 到点 C．那么首先使点跑过道路 AB 然后再跑过 BC，我们就得到道路 AC，自然把它称为道路 $U=AB$ 和 $V=BC$ 的乘积，并且表为 UV．如果道路 U，V 对应地等价于道路 U_1，V_1，那么它们的乘积 UV 和 U_1V_1 也将等价．

道路的乘法在下述意义下是结合的：乘积 $U(VW)$或$(UV)W$

中的一个是确定的，那么另外一个也就是确定的，而且两个是等价的道路．如果动点 M 以相反的方向跑过任意一条道路 $U=AB$，那就得到由点 B 到点 A 的与道路 $U=AB$ 相反的道路 $U^{-1}=BA$．道路 AB 和与它相反的道路 BA 的乘积是一条等价于仅由一个点 A 组成的零道路．

根据定义，不是任意两条道路都可以相乘，而仅当第一条道路的终点与第二条道路的起点相重合时才可相乘．这个缺点是可以消除的，如果我们仅考虑从同一个起点 A 出发的闭道路．任意两条这样的道路可以相乘，结果又得到一条带有起点 A 的闭道路．此外，对于每一条带有起点 A 的闭道路，它的逆道路具有同样的性质．

现在约定等价的道路认为是在曲面上仅以不同的方法所引出的同一条"道路"的不同的代表，而不等价的道路将认为是在本质上不同的道路的代表．以上所述指出，在这样的情况下从曲面的任意点 A 出发的所有闭道路的集合相对于道路的乘法运算成一群，零道路是这个群的单位元素，而已给的道路的逆元素仍是该路线，只是以相反的方向跑过．

一般讲，道路成的群不仅依赖于曲面的型而且依赖于起点 A 的选取．然而如果曲面不能分离为单独的块，即如果它的任意两个点都可以被一条位于曲面上的连续的道路所连结，那么由不同起点所得的道路群将是同构的，而且在这情况下可以简单地说关于曲面 S 的道路群，而不用指明点 A．曲面的这个道路群称为是它的**基本群**．

如果曲面 S 是平面或者球面，那么道路成的群仅由单位元素所组成，因为在平面和在球面上任意道路都可以缩为一点．然而如我们所看到的，在无穷圆柱面上有不能集中到一点的闭道路．既然在圆柱面上的由 A 点出发的任意闭道路，等价于道路 X 的某一次数(图 23)，而且不同次数的 X 彼此不等价，所以圆柱面的道路群是无限循回群．可以证明，环面(图 22)的道路群由型为 U^mV^n $(m, n=0, \pm1, \pm2, \cdots)$ 的道路组成，而且 $UV=VU$ 和

$U^mV^n=U^{m_1}V^{n_1}$ 仅当 $m=m_1$, $n=n_1$(提醒一下，在道路群中道路的等式了解为等价的意义).

道路群的重要性是因为它的以下性质. 我们假设，除曲面 S 外，给了另外一个曲面 S_1, 使得在曲面 S 和 S_1 之间可以建立双侧单值和连续的对应. 例如，这样的对应是可能的，如果曲面 S_1 是由借助于曲面 S 经过既不撕破又不把曲面不同的点粘起来的连续变形而得到. 对起始的曲面 S 上的每一条道路将在曲面 S_1 上对应一条道路. 在这情况下等价的道路将对应于等价的道路，两条道路的乘积对应于对应的道路的乘积，因而在曲面 S_1 上的道路群同构于在曲面 S 上的道路群. 换句话说，从抽象的观点来看，即在同构的意义下，道路群在曲面的所有可能的双侧单值连续的变换下是不变的. 如果两个曲面的道路群是不同的，那么所对应的曲面不可能连续地把一个变成另一个. 例如因为平面的道路群由一个元素组成，而圆柱面的道路群是无穷的，所以平面不可能不经粘合和撕破而变形成圆柱面.

图形在双侧单值和双侧连续的变换下保留不变的性质是在数学中一个特别的分支拓扑学中来研究，它们基本的思想在第十八章中阐明了. 连续变换的不变量称为是**拓扑不变量**. 道路群是拓扑不变量的重要例中的一个，很明显道路群不仅对曲面是可以定义的而且对任意点集也是可定义的，只要在这个集合里能说明什么是道路和它的变形.

定义关系 在拓扑里，道路群的计算方法已仔细地研究过. 并且发现，一般地说，这些群不得不用特殊的方法来决定，这个方法不仅对拓扑里的基本群可以用，在群论里它也常被用来给定抽象群. 下面就来叙述它.

设 G 是某个群. 元素 $g_1, g_2, \cdots, g_n$ 叫做群 G 的生成元，如果任意的元素 g 可以表示成形式

$$g=g_{i_1}^{\alpha_1}g_{i_2}^{\alpha_2}\cdots g_{i_k}^{\alpha_k},$$

此处 $i_1, i_2, \cdots, i_k$ 是数 1, 2, …, n 中的某一些；不相邻的指标 i 可以重复；因子的数目 k 任意；指数 $\alpha_1, \alpha_2, \cdots, \alpha_k$ 是不为零的正或

负整数.

为了充分地了解群 G，除了生成元以外，还要知道，怎样的乘积代表群的同一个元，怎样的乘积代表不同的元. 这样一来，对于群来说，必须列举在群 G 中成立的所有形如

$$g_{i_1}^{\alpha_1} g_{i_2}^{\alpha_2}\cdots g_{i_k}^{\alpha_k} = g_{j_1}^{\beta_1} g_{j_2}^{\beta_2}\cdots g_{j_l}^{\beta_l}$$

的等式. 由于这样的等式永远是无限多，所以通常不列举它们全体，而只给出这样的一些等式，由这些等式按照群的公理就能推出其他的. 这些等式就叫做定义关系.

显然，同一个群可以用不同的定义关系来给出.

例如，我们考虑具有生成元 a, b 和关系

$$a^2 = b^3,\ ab = ba \tag{10}$$

的群 H. 令 $c = ab^{-1}$，我们有

$$a = bc,\ a^2 = b^2c^2,\ b^3 = b^2c^2,\ b = c^2,\ a = c^3.$$

我们看到，群 H 的所有元素可以通过单一元素 c 来表达，而且

$$a = c^3,\ b = c^2.$$

因为关系式(10)可以直接由这些等式导出，所以对 c 没有任何非显然的关系. 因此群 H 是有生成元 c 的无限循回群.

如果在群里可以选取这样的一些生成元，它们没有任何非显然关系的约束，那么群叫做自由的，而所指出的生成元叫做自由生成元. 例如，要是群有自由生成元 a, b，那么它的每个元素可以唯一地写成形状

$$a^{\alpha_0} b^{\beta_1} a^{\alpha_1} b^{\beta_2} a^{\alpha_2} \cdots b^{\beta_k} a^{\alpha_k},$$

此处 $k = 0, 1, 2, \cdots, n$，指数 $\alpha_0, \beta_1, \alpha_1, \cdots, \beta_k, \alpha_k$ 表示不等于零的正或负整数，但"边上的" α_0 和 α_k 可以取零值. 类似的断言对于有更多数目的生成元的自由群也成立.

把两个没有公共元的群的生成元和定义关系写出，那么联合这些关系，我们得到新的群，叫做给定群的自由积.

自由群的理论，甚而自由积的一般理论在群论里都占有显明地位. 从几何观点看来，群 H_1 和 H_2 的自由积是这样的图形的

道路群，这个图形可以设想为只相交于一点的两个闭图形的和，这两个闭图形分别以 H_1 和 H_2 做为自己的道路群．我们已经知道，圆柱面的道路群是有一个生成元的自由群．由上面所做的说明，知道由图 24 所描写的曲面的道路群是有两个生成元的自由群．

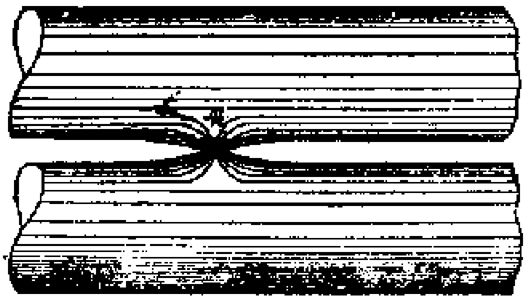


图 24

类似于曲面的基本群的定义，对空间的有限或无限体可以引进基本群．

扭结和扭结群 像已说的那样，从拓扑的观点看来，两个曲面认为是一样的，如果其中的一个可以用双侧单值和双侧连续的变换变成另一个．所有的闭曲面的拓扑分类问题早已解决．证明了每一个处于我们通常空间的闭曲面或者和球面拓扑等价，或者和具有某些个柄的球面(图 25)拓扑等价．例如，图 22 所描绘的环面可以连续形变为有一个柄的球面，立方体的表面可以连续形变为球面等等．由这个讨论，知道基本群的研究对闭曲面不很有意义，因为闭曲面的完全分类可以不用这些群．但是存在着不用基本群至今仍几乎不能成功从事的很简单的例．著名的扭结问题便属于这类．

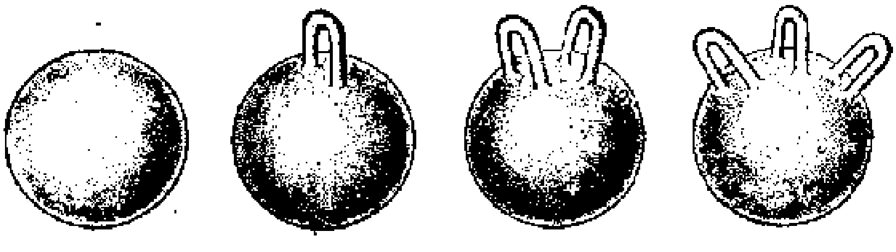


图 25

我们把安置在通常三维空间中的闭曲线叫做扭结．这个安置像图 26 所显示的那样可以很不同．两个扭结叫做等价的，如果其中的一个可以连续的形变为另一个，但在形变过程中既不撕断曲线又没有地方重叠．立刻发生两个问题：1) 怎样知道由平面图

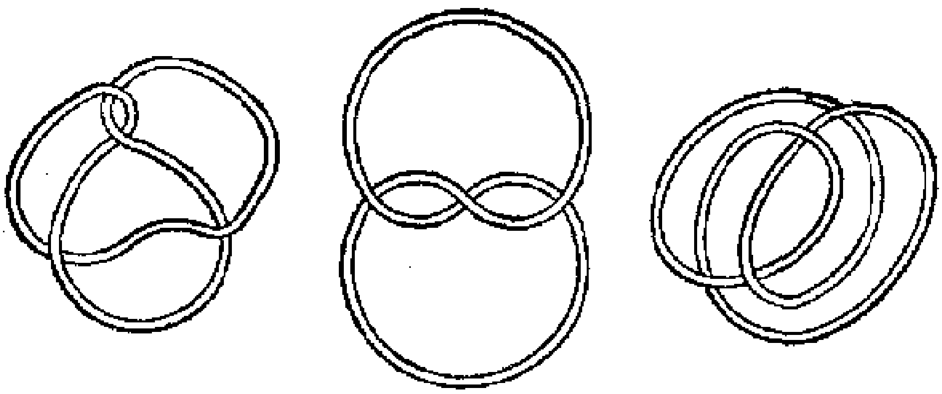


图 26

形给出的任意两个扭结是否等价；2) 所有的不等价扭结怎样分类。

两个问题至今仍未解决，并且在它们的部分解中目前有的基本成果是和群论联系着的。

我们把空间中属于给定扭结的点取去，并且考虑余下点集的基本群．这个群也叫做扭结群．直接可以知道，如果扭结等价，那么它们的群同构．因此由扭结群的不同构可以做出扭结本身不等价的结论．例如，圆周的扭结群是循环群，而有三叶玫瑰曲线形状（图 27）的扭结群是一个更复杂的群．由于后者是不可交换的，所以和圆周的群不同构．因此可以断定三叶玫瑰形扭结不切开就不能展直成圆周．这个事实试验起来是显然的，但做为证明需要精细的数学理由．

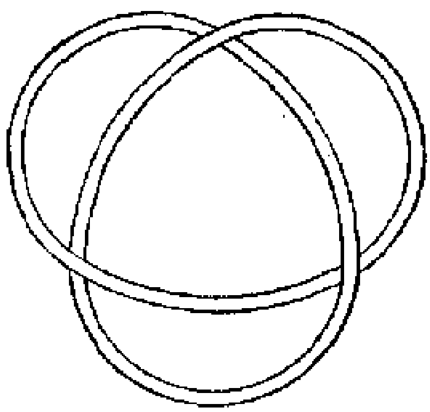


图 27

可惜，在考虑扭结群时，也出现一些至今没有解决的困难问题．问题是，在拓扑中虽有很简单的办法可以按已给的扭结图象找出扭结群的生成元和定义关系，但是为了要利用这些群来比较不同的扭结，还需要解决由自己的生成元和定义关系所确定的群是否同构，而这个问题的解法到现在还不知道．不但如此，苏联数学家诺维科夫不久前证明了一个卓越的定理：不可能指出任何的

唯一正则方法(精确点讲是所谓正规算法)，利用它总可以解决具有相同生成元的两组已给定义关系是否确定同一个群．这个定理使得我们对下述的那样一个一般方法的存在性产生怀疑，这个方法能用来判别由平面图象所给出的扭结的等价性．

§9．群的表示与指标(特征标)

群的一般理论就其方法来说有一点像初等几何学：两者都是以确定的公理系统作为基础，并由此出发建立起理论的全部内容．然而，解析几何的例子告诉我们，为了研究几何问题，引进分析的数值方法是非常有利的．

分析与古典代数的方法对群论的应用被称之为群表示论．就像解析几何不仅是借分析的帮助，给与解决几何问题的方法，而且反转过来，在许多复杂的分析问题上投下了几何的光辉一样，群表示论在更大程度上也是如此．它不仅供给了研究群的性质的辅助工具，而且把分析与群论的一些概念密切地联系起来．它不仅对于群论的一些事实找到了用数值关系的表示法，而且对于分析的一些关系，也找到了群的解释．现在，群论在物理上的很大一部分的重要应用是与表示论密切联系着的．

群的矩阵表示　在线性代数中(参看第十六章)已经讨论过矩阵乘法的运算．这个运算是结合的，但一般地说，是非交换的．阶数给定的全体非退化(亦称非奇异)方阵对乘法来说做成一个群．事实上，任意两个非退化方阵的乘积仍然是一个非退化方阵，单位方阵起着群的单位元的作用，并且对于每一非退化方阵，存在其逆方阵，也是非退化的．

设给出某个群 G，并且对于 G 的每一元素 g，有一确定的复数的 n 阶非退化方阵 A_g 与之对应，而且，对于群的元素的乘积，也对应着与它们对应的矩阵的乘积：$A_{gh}=A_g\cdot A_h$．这时我们就说，给出了群 G 的一个 n 阶矩阵表示．通常“矩阵”这个字可以省略，而简单地说群 G 的一个 n 阶表示．群 G 的一个 n 阶表示是 G 到

n 阶非退化矩阵群里的一个同态映象. 由同态映象的一般性质可知,对于任意表示,群 G 的单位元的对象是单位矩阵, 在 G 中互为逆元的象互为逆矩阵.

由于一阶方阵是单独的复数, 故群 G 的一阶表示是这样的对应: 对于 G 的每一元素,对应着一个复数, 而且, 群的两个元素的乘积对应着与它们对应的复数的乘积. 例如, 对于对称群的偶置换来说,命数 1 与之对应,而对于奇置换,命数 -1 与之对应,则这个对应便是对称群的一个 1 阶表示.

对于群 G 的每一个元素,命 n 阶单位方阵 E 与其对应,这样,我们便得到了群 G 的一个表示, 称之为 n 阶单位表示. 如果 G 是有限群并且含有非单位元素, 那么, 除了不同阶数的单位表示之外, 群 G 必然还有无限多个其他表示. 以下将指出求这些表示的方法.

设已经知道群 G 的一个表示, 则可得出无限多个其他表示. 事实上, 设 $g\to A_g$ 是群 G 的给定的一个 n 阶表示, 任取一个 n 阶非退化矩阵 P, 命 $B_g=P^{-1}A_gP$, 则 $g\to B_g$ 仍是群 G 的一个表示, 因为

$$B_{gh}=P^{-1}A_{gh}P=P^{-1}A_gA_hP=P^{-1}A_gPP^{-1}A_hP=B_gB_h.$$

由一已知表示出发,选择各种不同的矩阵 P, 可得出许多表示, 这样得出的表示叫做与已知表示**等价的**. 在表示论里, 等价的表示本质上不认为是不同的,被讨论的所有表示通常仅精确到等价性.

求新表示的另外方法是作表示的直接和,此方法可述之如下: 设 $g\to A_g$, $g\to B_g$ 是群 G 的矩阵的某两个表示,相应的阶数分别为 m 与 n. 考虑对应

$$g\to\begin{pmatrix}A_g & O\\ O & B_g\end{pmatrix}.$$

回忆矩阵的乘法规则(参看第十六章),我们得到

$$gh\to\begin{pmatrix}A_{gh} & O\\ O & B_{gh}\end{pmatrix}=\begin{pmatrix}A_gA_h & O\\ O & B_gB_h\end{pmatrix}=\begin{pmatrix}A_g & O\\ O & B_g\end{pmatrix}\begin{pmatrix}A_h & O\\ O & B_h\end{pmatrix},$$

亦即上述映象仍是群 G 的一个表示. 这个表示叫做已知的两个表

示的和，而用符号 A_g+B_g 表之．如果交换被加项的位置，那么，我们得到另一个表示

$$g\to\begin{pmatrix}B_g & O\\ O & A_g\end{pmatrix}.$$

但是，这个表示是与原先得出的表示等价的．因此，如果对于等价的表示不加区别，那么，表示的加法是可换的运算．很容易看出，表示的加法也是结合的运算．设已经知道了群 G 的某些表示 A_g, B_g, C_g, …，那么，利用表示的加法可以得出更高阶的表示：$A_g+B_g+C_g$, $A_g+A_g+A_g+A_g$ 等等．

例如，数 1, −1, i, $-i$ 关于乘法组成一个群．对于这个群的每一个数，命这个数本身和它对应，这样我们便得到一个一阶表示．其次，我们取映象 $1\to1$, $-1\to-1$, $i\to-i$, $-i\to i$, 作为第二个表示．这两个表示的和是映象

$$1\to\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad -1\to\begin{pmatrix}-1&0\\0&-1\end{pmatrix},\quad i\to\begin{pmatrix}i&0\\0&i\end{pmatrix},\quad -i\to\begin{pmatrix}-i&0\\0&i\end{pmatrix}.$$

利用矩阵 $P=\begin{pmatrix}1&i\\1&-i\end{pmatrix}$ 来变换这个表示，我们得到等价的表示

$$1\to\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad -1\to\begin{pmatrix}-1&0\\0&-1\end{pmatrix},$$

$$i\to\begin{pmatrix}0&1\\-1&0\end{pmatrix},\quad -i\to\begin{pmatrix}0&-1\\1&0\end{pmatrix}.$$

让我们很有兴趣地指出，这个表示的所有矩阵都是实矩阵．

设群 G 的某个 n 阶表示的所有矩阵都有以下形式：

$$g\to A_g=\begin{pmatrix}B_g & C_g\\ 0 & D_g\end{pmatrix},$$

此处 B_g, D_g 是方阵，而 A_g 的左下角的元素均为零．矩阵 A_g 与 A_h 相乘，我们得到

$$A_{gh}=A_gA_h=\begin{pmatrix}B_gB_h & B_gC_h+C_gD_h\\ 0 & D_gD_h\end{pmatrix},$$

即 $B_{gh}=B_gB_h$, $D_{gh}=D_gD_h$．这个事实表明，映象 $g\to B_g$, $g\to D_g$ 也

是群 G 的表示，但阶数较低．表示 A_g 叫做群 G 的阶梯表示，而每一个与其等价的表示均称之为可约的，与任何阶梯表示不等价的表示叫做不可约的．

如果所有矩阵 A_g 不但左下角全部为零，而且右上角的 C_g 也全是零，那么就说表示 A_g 可分解为表示 B_g, D_g 的和．与不可约表示的和等价的表示叫做完全可约的．

在群论中证明了以下事实：有限群的每一表示都是完全可约的[1]．由此可知，为了求有限群的所有表示，只要知道它的所有不可约表示即可，因为所有其余表示都与不可约表示的各种和等价．

任何一个有限群的不可约表示的实际计算通常都是相当复杂的问题，仅对于个别的有限群的类，问题被用明显的形式解决了，例如，对于可换群、对于对称群以及某些其他的群，虽然从理论观点来看，有限群的表示的性质是相当仔细地被研究了．

对于每一个有限群，都有用以下方法引入的特别"正则"表示：设 g_1, $g_2\cdots$, g_n 是给定群 G 的按照任意顺序排起来的全部元素，并设

$$g_ig_k=g_{i_k}\quad (i, k=1, 2, \cdots, n).$$

对 k 选择某个固定的值，作一个 n 阶方阵，使其第 i 行第 i_k 个位置为 1，而其余地方均为零($i=1, 2, \cdots, n$)，并且用符号 Rg_k 表示这个方阵．通过简单的计算，很容易证明对应 $g_k\to Rg_k$($k=1, 2, \cdots, n$)的确是一个表示，称这个表示为群 G 的正则表示．

也可以证明：当群的元素的编号改变时，我们得出的表示是与上述表示等价的．因此，就精确到等价的意义来说，每一个有限群仅有一个正则表示．

让我们简短地叙述一下有限群表示论的一些基本定理．有限群的不同的(非等价的)不可约表示的个数是有限的，并且等于这个群的共轭元素类(参看 285 页)的个数．不可约表示的阶数必然是群的阶数的因数，而且正则表示等于所有不可约表示的和，其中

1) 让我们回忆起：所讨论的群的表示，矩阵的元素可以是任意复数．

每一个不可约表示重复出现的次数恰好等于其阶数.

由此可以得出，有限群的阶数与不可约表示阶数之间的下面有趣的关系式.

设群 G 的元素的个数用 n 表示，共轭元素类的个数用 k 表示，而 G 的不可约表示的阶数分别用 $n_1, n_2, \cdots, n_k$ 表示. 由正则表示的作法，知其阶数是 n，而正则表示等于 n_1 个第一个不可约表示的和，加 n_2 个第二个不可约表示的和，再加上 n_3 个第三个不可约表示的和等等，一直加到 n_k 个第 k 个不可约表示的和，故有以下等式：

$$n=n_1^2+n_2^2+\cdots+n_k^2. \qquad (11)$$

让群的每一个元素与数目 1 对应，于是，我们得到一个阶数为 1 的不可约表示，这是每一个群都具有的显然表示. 设在式子 (11) 中的 n_1 就是这个单位表示的阶数，于是，(11) 可写成

$$n=1+n_2^2+\cdots+n_k^2$$

的形式，这时 $n_2, \cdots, n_k$ 表示不可约表示的阶数.

利用 $n_2, \cdots, n_k$ 应该是 n 的因数这个事实，并且知道了 k，那么仅由等式 (11) 就可求出 $n_2, n_3, \cdots, n_k$. 例如，三个元素的置换对称群 S_3 有三个共轭元素类：(1)；(12)，(13)，(23)；(123)，(132). 此处 $n=6$ 而 $k=3$，等式 (11) 有唯一的一组解：$6=1^2+1^2+2^2$. 因此 S_3 由两个不同的一阶表示与一个不可约的二阶表示组成.

有限可换群可作为另外一个例子，这时每一个元素组成一个共轭元素类. 因此，$k=n$，由式子 (11) 可知 $n_1=n_2=\cdots=n_k=1$，即这样群的所有不可约表示都是一阶的，而且不可约表示的个数等于群的阶数.

可换群的不可约表示也叫做群的指标. 对于非可换群的每一表示，其指标是指组成这个表示的矩阵的迹(即矩阵对角元素的和)的集合. 有限群的指标具有很好的性质和关系. 群的表示和指标的研究以有兴趣的一般结果丰富了群论，并且这些结果在近代理论物理中找到了宽广的应用.

§10. 一 般 群 论

我们已经指出，几乎在上世纪的整个时期中，群论是主要作为变换群的理论而发展的．然而，逐渐看到，群本身的研究是主要的，而变换群的研究可归结为抽象群及其子群的研究．

由变换群论向抽象群论的过渡开始是在有限群论中来完成的．李群的迅速发展，以及群论之深入拓扑学，都要求建立一般的群论，而把有限群仅作为其某个特殊情形来讨论．

1916 年在基辅出版的施密特所写的关于群论的书，是非常清晰地表明这个观点的第一本教科书．在 20 年代，施密特得到了关于无限群理论的一些重要定理，成为一些苏维埃代数学家研究的出发点．由于施密特及 H. C. 亚历山大洛夫为普及近世代数概念所做工作的结果，在莫斯科形成了群论学派，他们的学生库洛什很快就成为这个学派的领导人．特别是由他证明的关于自由积的每一子群本身是分别与乘积因子的子群同构的子群的自由积的定理而著名．后来，他写了一本关于群论的专门著作，这是头一本系统地叙述在一般群论的领域中已得到的丰富成果的书．在现在，这是一本在全世界范围内最完备的群的一般理论教程，并且享有国际声誉．

继莫斯科代数学家之后，列宁格勒以及其他城市的代数学家也从事一般群论的研究，对群论的发展作出了很大的贡献．现时，对群论的研究，包括其所有主要部分，苏联数学家起着主要作用，得出一些结果不止一次地表明在群论的发展上起着决定性的影响．

§11. 超 复 数

应用代数方法解决实际问题时，在比较简单的情形，通常归结为一个或若干个方程，其中以所要求的值作为其未知量．这时未

知量是所研究对象的数量描述；而方程的布列是借分析对象间存在着的实际关系而得出.

当所讨论的最简单的量是类似于质量、体积、长度的这些情形，为了描述它们的数量，用一个数即已足够．然而，在一些具体问题中，常遇到不仅用一个数描述的对象．恰恰相反，随着技术的发展，具有较复杂性质的对象，常具有较大的价值，而描述它们，必须要用到几个数甚至是无限多个数．例如，像力、速度、加速度这样一些重要的物理量，描述它们，就需要三个数．其次，我们都知道，描述空间中点的位置，需要用三个数，描述空间中平面的位置，也要用三个数，而描述空间中直线的位置，需要用四个数，而刚体的位置，甚至需要用六个数来描述．因此，在利用代数方法解决涉及较复杂对象的问题时，常会得到多个未知量的方程．而解这样的方程，常常是比利用这个对象的几何性质或物理性质直接解答困难得多．由此自然发生这样的思想：企图描述较复杂的对象，不利用一组通常的数，而是利用某种较复杂的广义的数，对于这种广义的数，也可以施行与通常算术运算类似的运算．问题的这样提法是比较自然的，科学的历史告诉我们，数的概念不是不变的，而其变化，都是逐次扩张数的集合．由自然数扩张至分数，然后由分数扩张至正负数、至实数（有理数与无理数），最后，至复数.

复数　读者由第四章（第一卷）已经知道复数的基本性质及其简单应用．现在，我们的兴趣仅在于复数概念的基础．让我们回忆一下复数概念的本身通常是如何定义的．开始仅讨论通常的实数，并指出负数的平方根没有意义，因为实数的平方是正数或者零．进一步指出，科学的迫切需要，使得数学家作为一种特殊种类的数来讨论表示式 $a+b\sqrt{-1}$，把它称之为虚数，以区别于通常的实数．如果认定这些虚数也适合通常的数所适合的算术运算法则，则所有负数的平方根都可用量 $i=\sqrt{-1}$ 来表示，而对实数和虚数进行有限次算术运算的结果，永远可表示 $a+bi$ 的形状，此处 a, b 是实数.

显然，虚数这样的定义在颇大程度上是与正常的思维矛盾的，

因为开始断言表示式$\sqrt{-1}$, $\sqrt{-2}$等等没有意义，然后又称这些没有意义的表示式为虚数. 这种情况使得十七世纪和十八世纪的数学家对使用复数的合理性引起极大怀疑. 但是，这种怀疑在十九世纪初便消失了，因为这时发现了复数用平面上的点的几何表示. 稍后不久，匈牙利数学家伯依阿依和英国数学家汉弥登给出了复数理论严格的纯算术的基础，这个基础叙述在下面.

讨论实数对(a, b)以代替数$a+bi$. 两个数对当它们的第一个位置的数和第二个位置的数分别相等时，称之为相等的数对，即当且仅当$a=c$, $b=d$时，认为$(a, b)=(c, d)$. 数对的加法和乘法用下面式子来定义：

$$(a, b)+(c, d)=(a+c, b+d);$$
$$(a, b)\cdot(c, d)=(ac-bd, ad+bc).$$

例如，我们有

$$(2, 3)+(1, -2)=(3, 1), (2, 3)\cdot(1, -2)=(8, -1),$$
$$(3, 0)+(2, 0)=(5, 0), (3, 0)\cdot(2, 0)=(6, 0).$$

后面的例子告诉我们，特别当数对的第二个位置的数是零时，这时数对的运算可归结为第一个位置的数的运算，因此，这样的数对可简单的用其第一个位置的数来表示. 其次，对于数对$(0, 1)$，我们用符号i来表示，于是，有

$$(a, b)=a(1, 0)+b(0, 1)=a+bi,$$
$$i^2=(0, 1)(0, 1)=(-1, 0)=-1,$$

这样，便得出了复数的通常表示法.

因此，从叙述的观点来看，复数是普通实数的数对，复数的运算不过是特殊种类的实数对的运算.

超复数 复数的多种多样有成效的应用，促使数学家在十九世纪头几十年就考虑这样问题：是否可以像由实数对建立复数那样，由三个实数的数组、四个实数的数组等等建立超复数. 从上一世纪的中叶开始，各种各样不同的特殊的这类超复数被研究着，而在上一世纪末以及本世纪初研究了超复数的一般理论，并且发现它在数学和物理的相间领域中有着许多的重要应用.

这样，称 n 个实数的数组 $(a_1, a_2, \cdots, a_n)$ 为 n 阶超复数，其中的每一个实数，称之为这个超复数的坐标．两个超复数 $(a_1, a_2, \cdots, a_n)$ 与 $(b_1, b_2, \cdots, b_n)$ 当其对应坐标相等，即 $a_1=b_1$, $a_2=b_2$, $\cdots$, $a_n=b_n$ 时叫做相等的．加法运算自然的被类似于复数的加法公式所定义，即

$$(a_1, a_2, \cdots, a_n)+(b_1, b_2, \cdots, b_n)$$
$$=(a_1+b_1, a_2+b_2, \cdots, a_n+b_n).$$

很自然地引入实数对超复数的乘法：规定

$$a(a_1, a_2, \cdots, a_n)=(aa_1, aa_2, \cdots, aa_n).$$

除此以外，应规定两个超复数的乘法，而且，其结果应该是一个超复数．

把通常复数的乘法定义推广到一般情形是很困难的．乘法可以用各种不同的方法来定义，因而将得出各种不同的超复数系．因此，首先应该弄清楚，这样的定义应该达到什么目的．无疑的，希望所定义出的超复数的运算将具有通常实数运算所具有的性质．然而，这些通常运算具有哪些性质？

仔细地考查数及其运算在代数上最常利用的性质，很容易发现，可归结为以下几条：

1．对于任意两个数，它们的和是唯一确定的．

2．对于任意两个数，它们的积是唯一确定的．

3．存在着一个数零，它具有性质：对于任意 a，均有 $a+0=a$，

4．对于每一个数 a，均存在其负数 x，适合等式 $a+x=0$．

5．加法适合交换律

$$a+b=b+a.$$

6．加法适合结合律

$$(a+b)+c=a+(b+c).$$

7．乘法适合交换律

$$ab=ba.$$

8．乘法适合结合律

$$(ab)\cdot c=a\cdot(bc).$$

9．乘法对加法适合分配律

$$a(b+c)=ab+ac,\ (b+c)a=ba+ca.$$

10．对于每一 a 以及每一 $b\neq 0$，存在唯一的数 x，满足等式 $bx=a$.

仔细分析的结果，得出了上面性质 1—10．近百年来数学的发展指出这些性质的极大重要性．现在，每一个满足性质 1—10 的量的系统，称之为域．例如，全体有理数的集合、全体实数的集合、全体复数的集合都是域，因为，对于这些集合中的每一个来说，任意两个数的相加、相乘都还在这个集合内，并且具有性质 1—10．除了这三个最重要的域外，可以举出数所做成的无限多个域．然而，除了数目做成的域以外，对于其他性质的量所做成的域，我们也有很大的兴趣．例如，在中学中，我们曾对被称之为代数分式的东西进行过运算，所谓代数分式，即这样的分式，其分子分母都是某些字母的多项式．代数分式可以相加、相减、相乘、相除，而且这些运算具有性质 1—10．因之，代数分式所组成的元素的系统是一个域．可以举出各种各样的由具有复杂性质的量所组成的域的例子．由于作为域的定义的性质 1—10 的重要性，首先，我们将提出这样问题，求出超复数这样的乘法运算，使得超复数作成一个域．如果这样问题能够顺利解答，那么，我们将得到新的、更一般的复数．然而，在上世纪一开始，就已发现仅对于阶数 2 的超复数，这个问题才是可能的，而且得出的仅是通常的复数．这个结果指出，复数占有完善的、特殊的地位．如果不改变性质 1—10 的要求，希望得到一个广义的数系，超过复数范围是不可能的．因此，为了进一步尝试建立新的数系，必须放弃性质 1—10 中的一个或者某几个．

四元数　在数学上被讨论的第一个超复数系是四元数系，是上一世纪中叶英国数学力学家汉弥登引入的．这个系统满足性质 1—10 中除 7(乘法交换律)以外的所有要求．

四元数可以用下面方式来叙述．对于四个数的数组(1, 0,

0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1) 引入简单的记号 1, i, j, k. 于是，由于等式

$$(a, b, c, d)=a(1, 0, 0, 0)+b(0, 1, 0, 0)+c(0, 0, 1, 0)+d(0, 0, 0, 1)$$

成立，每一个四元数可唯一的表成以下形式：

$$(a, b, c, d)=a\cdot 1+b\cdot i+c\cdot j+d\cdot k.$$

四元数 1 将认为是这个系统的单位量，即认为对于任意四元数 α，有 $1\cdot\alpha=\alpha\cdot 1=\alpha$. 其次，定义 $i^2=j^2=k^2=-1$,

$$ij=-ji=k,$$
$$ik=-ki=-j,$$
$$jk=-kj=i.$$

利用图 28，很容易记住四元数的这个"乘法表". 在图形中，圆上的点 i, j, k 分别表示四元数 i, j, k. 这三个四元数中任意两个相邻的乘积等于第三个，当第一个因子到第二个因子是按照顺时针方向旋转时，作为乘积的第三个四元数取正号，反之，则取负号. 知道了四元数 i, j, k 的乘法表以后，利用乘法对加法的分配律 9，就可求出任意两个四元数的乘积，即

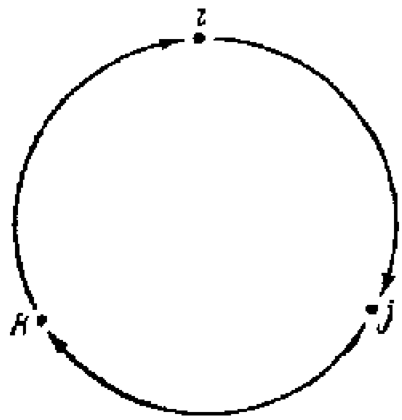


图 28

$$\begin{aligned}(a\cdot 1+b\cdot i+c\cdot j+d\cdot k)(a_1\cdot 1+b_1\cdot i+c_1\cdot j+d_1\cdot k)\\ =aa_1\cdot 1+ab_1\cdot i+ac_1\cdot j+ad_1\cdot k+ba_1\cdot i+bb_1\cdot ii+bc_1\cdot ij\\ +bd_1\cdot ik+ca_1\cdot j+cb_1\cdot ji+cc_1\cdot jj+cd_1\cdot jk+da_1\cdot k\\ +db_1\cdot ki+dc_1\cdot kj+dd_1\cdot kk\\ =(aa_1-bb_1-cc_1-dd_1)\cdot 1+(ab_1+ba_1+cd_1-dc_1)\cdot i\\ +(ac_1+ca_1-bd_1+db_1)\cdot j+(ad_1-da_1+bc_1-cb_1)\cdot k.\end{aligned}$$

四元数中第一项的因子 1 通常被省略，而把 $a\cdot 1$ 写成 a. 等式 $ij=-ji$, $ik=-ki$, $jk=-kj$ 表明四元数的乘法不适合交换律. 这里乘数与被乘数的地位不是平等的，因此，在计算四元数时必须

细心地注意因子的顺序. 除此之外,对四元数进行运算,没有任何困难. 特别,四元数乘法也适合结合律8. 这个事实很容易借乘法表对基底四元数1, i, j, k来验证;而由此过渡到一般情形是很明显的.

数a叫做四元数$a+bi+cj+dk$的实数部分,或者纯量部分,而把和$bi+cj+dk$叫做$a+bi+cj+dk$的向量部分. 向量部分仅相差一个符号的两个四元数$a+bi+cj+dk$与$a-bi-cj-dk$叫做共轭的. 显然,两个共轭的四元数的和是一个实数. 此外,按照上面写出的公式求两个共轭四元数的乘积,得到

$$(a+bi+cj+dk)(a-bi-cj-dk)=a^2+b^2+c^2+d^2, \quad (12)$$

即两个共轭的四元数的乘积也是一个实数.

四元数$a+bi+cj+dk$的系数的平方和$a^2+b^2+c^2+d^2$叫做这个四元数的范数. 由于任一实数的平方都是非负的数,故任一四元数的范数也是非负的数,仅当这个四元数是0时,范数才能是零.

公式(12)表明, 任一四元数与其共轭的乘积等于该四元数的范数.

我们约定用星号表示一个四元数的共轭, 即用α^*表四元数α的共轭. 于是,直接计算可证明以下公式是正确的:

$$(\alpha\beta)^*=\beta^*\alpha^*.$$

由此得出一个有趣的推论: 四元数乘积的范数等于因子范数的乘积. 事实上,根据上面式子,我们有

$$\text{范数}(\alpha\beta)=(\alpha\beta)(\alpha\beta)^*=\alpha\beta\beta^*\alpha^*=(\alpha\alpha^*)(\beta\beta^*)$$
$$=\text{范数}\,\alpha\cdot\text{范数}\,\beta.$$

由于范数的性质, 我们可以非常简单地解决关于四元数相除问题. 设$\alpha=a+bi+cj+dk$是任意一个非零四元数,于是

$$(a+bi+cj+dk)\frac{1}{a^2+b^2+c^2+d^2}(a-bi-cj-dk)$$

$$=\frac{1}{a^2+b^2+c^2+d^2}(a^2+b^2+c^2+d^2)=1,$$

即四元数

$$\frac{1}{a^2+b^2+c^2+d^2}(a-bi-cj-dk)=\alpha^{-1}$$

是给定的四元数 α 的逆元.

会求四元数的逆元以后,很容易就可求出两个四元数的商.事实上,设给出两个四元数 α, β, 且第一个四元数不是零四元数. 为了求 β 被 α 除的商,我们应该解下面两个方程

$$\alpha x=\beta,\ y\alpha=\beta.$$

第一个方程的两端同用 α^{-1} 左乘,我们得

$$x=\alpha^{-1}\beta.$$

第二个方程的两端同用 α^{-1} 右乘,我们得

$$y=\beta\alpha^{-1}.$$

因为 $\alpha^{-1}\beta$ 与 $\beta\alpha^{-1}$ 一般情形是不同的,故对于四元数来说,必须区分开两种除法: 右除和左除. 两者永远是可施行的,当然,以零为除数的应除外.

向量代数 虽然四元数的运算在很多方面与复数的运算相类似,但是,由于乘法不适合交换律,使得四元数的性质与数目的性质有很大差异. 例如,由复数的代数,我们都知道,二次方程有两个根. 如果我们在四元数的范围内,即使讨论二次方程

$$x^2+1=0,$$

那么,可以看出以下的六个四元数都是它的根: $\pm i$, $\pm j$, $\pm k$. 更精确的分析,可以发现它有无限多个不同的解. 这种情况使得四元数在数学上的应用带来很大困难,尽管汉弥登以及另外一些数学家把四元数引向数学与物理的各个部门作了很多的尝试,但到现在为止,四元数在数学上所起的作用仍然是比较小的,无论如何也不能与复数的作用相比较.

然而,四元数给与了作为近代技术与物理的不可缺少的工具向量代数的发展以很大的推动力. 这是由于在力学和物理中,速度、加速度、力等概念起着重要的作用,为了描述它们,需要用三个数. 上面我们已经看到,每一个四元数可以看作实数 a 与向量部

分 $bi+cj+dk$ 的集合. 由于向量部分是被三个数所确定的，由此，为了描述重要的物理量，利用四元数的向量部分即已足够.

可以利用从直角坐标系原点引出、在三个坐标轴上的射影分别为 b, c, d 的向量来几何地表示四元数 $a+bi+cj+dk$ 的向量部分. 因此，任意一个四元数可以用几何方法表成空间中的向量与数的总和. 让我们来考查一下四元数的运算可以有怎样的解释.

取纯量部分等于零的两个四元数（向量四元数）$xi+yj+zk$ 与 $x_1i+y_1j+z_1k$，用几何方法把它们用从坐标原点引出的两个向量表出. 这两个四元数的和仍是向量四元数 $(x+x_1)i+(y+y_1)j+(z+z_1)k$. 很容易看出，表示这个和的向量是头两个向量所构成的平行四边形的对角线. 因此，向量四元数的加法与非常熟悉的按平行四边形规则的向量加法运算相对应. 类似地，如果用某个实数乘向量四元数，那么，表示这个四元数的向量也应用这个数去乘.

当两个向量相乘时，我们将遇到另一种情况. 实际上，

$$(xi+yj+zk)(x_1i+y_1j+z_1k)$$
$$=-xx_1-yy_1-zz_1+(yz_1-y_1z)i+(zx_1-z_1x)j$$
$$+(xy_1-x_1y)k,$$

也就是说，两个向量四元数部分相乘，我们得一个完全的四元数，既有纯量部分，也有向量部分.

向量四元数的乘积的纯量部分，并取反号，叫做表示该四元数的向量的纯量积，而乘积的向量部分叫做表示该四元数的向量的矢量积. 向量 α, β 的纯量积通常用符号 (α, β) 表示，或者简单地表为 $\alpha\beta$，而这两个向量的矢量积用符号 $[\alpha, \beta]$ 表示. 设 $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ 分别是表示四元数 i, j, k 的向量，即在相应的坐标轴上的单位长的向量. 按照定义，如果 $\alpha=x\boldsymbol{i}+y\boldsymbol{j}+z\boldsymbol{k}$, $\beta=x_1\boldsymbol{i}+y_1\boldsymbol{j}+z_1\boldsymbol{k}$，则

$$(\alpha, \beta)=xx_1+yy_1+zz_1,$$

$$[\alpha, \beta]=(yz_1-y_1z)\boldsymbol{i}+(zx_1-z_1x)\boldsymbol{j}+(xy_1-x_1y)\boldsymbol{k}.$$

利用上面的式子，容易给与向量的纯量积与矢量积以几何解释. 这就是两个向量的纯量积等于这两个向量的长与其夹角的余弦的乘积，而两个向量的矢量积是一个向量，其长等于该两个向量

所构成的平行四边形的面积，且它垂直于上述平行四边形所在平面．而从它所在的一侧来看，由给出的第一个向量到第二个向量的旋转的转向正好像从 Oz 轴所在的一侧看 Ox 轴到 Oy 轴的旋转一样．

现时在力学和物理中，照例不应用四元数的运算，而仅讨论向量的运算以代替它，并且这些运算按照刚才所叙述的结果用纯几何方法来定义．

最后，让我们举出借四元数可以非常漂亮的解答的一个力学问题．它的解答是发现四元数的原因之一．

设刚体先围绕通过已知点 O 的确定的轴 OA 按照给定的方向旋转某个角 φ，然后再围绕通过 O 的另一轴 OB 旋转角 φ_1，求这个刚体围绕哪一个轴、旋转怎样的角度直接从第一个位置旋转到第三个位置？这是力学上关于有限个旋转复合的著名问题．诚然，它可以利用通常的解析几何来解决，在十八世纪欧拉就曾经这样做过．但是，当利用四元数时，它的解答可以有非常简明的形式．

设 $\xi=xi+yj+zk$ 和 $\alpha=a+bi+cj+dk$ 是两个四元数，其中第一个我们将认为是变动的，而第二个是给定的．经过计算，很容易验证表示式 $\alpha^{-1}\xi\alpha$ 是向量四元数．现在，如果用向量 $\boldsymbol{\xi}$, $\boldsymbol{\xi}_1$, $\boldsymbol{\alpha}$ 表示四元数 ξ, $\alpha^{-1}\xi\alpha$ 以及四元数 α 的向量部分，那么，向量 $\boldsymbol{\xi}_1$ 可由向量 $\boldsymbol{\xi}$ 围绕通过向量 $\boldsymbol{\alpha}$ 的轴旋转 φ 角而得出，此处角 φ 被式子 $\cos\dfrac{\varphi}{2}=\dfrac{a}{\sqrt{a^2+b^2+c^2+d^2}}$ 所决定．因此，可以认为四元数 $\alpha=a+bi+cj+dk$ 表示围绕轴 $\boldsymbol{\alpha}=b\boldsymbol{i}+c\boldsymbol{j}+d\boldsymbol{k}$ 作角 φ 的空间旋转．

反之，知道了旋转轴和角度 φ 以后，可以求出表示这个旋转的四元数．这样的四元数有无限多个，但是，它们彼此间仅相差一个常数因子．

现在，让我们再考虑一个旋转，角度为 φ_1 旋转轴为 $\boldsymbol{\beta}=b_1\boldsymbol{i}+c_1\boldsymbol{j}+d_1\boldsymbol{k}$．设这个旋转用四元数 $\beta=a_1+b_1i+c_1j+d_1k$ 表示．设任意向量 $\boldsymbol{\xi}$ 在第一个旋转作用下得到向量 $\alpha^{-1}\boldsymbol{\xi}\alpha$．得到的这个向量再经第二个旋转作用，得到 $\beta^{-1}(\alpha^{-1}\boldsymbol{\xi}\alpha)\beta$．根据结合律，最后的

结果可以表为下述形式：

$$\beta^{-1}(\alpha^{-1}\xi\alpha)\beta=(\alpha\beta)^{-1}\xi(\alpha\beta).$$

由于向量四元数 ξ 左乘四元数 $(\alpha\beta)^{-1}$ 并且右乘四元数 $\alpha\beta$ 的乘积等于向量 ξ 围绕相应轴作相应角度的旋转，故我们得出这样的结论：依次施行与四元数 α, β 相应旋转的结果是与乘积 $\alpha\beta$ 相应的旋转．换句话说，旋转的加法对应于这些旋转的四元数的相乘．

四元数除了在几何和物理上的应用外，在数论上也有极好的应用．在这个领域中应特别指出林尼克的工作．

§12．结 合 代 数

代数(超复数系)的一般定义 前面为超复数所下定义中必须用到实数，把超复数简单地看作实数组．然而，这样的观点是太狭窄了，为了理论研究的需要，逐渐地采取下面一般的定义．

某个量的系统 S 叫做域 P 上的代数(或超复数系)，如果

甲) 对于域 P 中每一元素 a 以及系统 S 中每一个量 α，确定这个系统中唯一的元素，称之为 a 与 α 的乘积，并用记号 $a\alpha$ 表示；

乙) 对于系统 S 中每两个量 α, β，唯一确定这个系统中某一个量，称之为 α, β 的和，并用记号 $\alpha+\beta$ 表示；

丙) 对于系统 S 中每两个量 α, β，唯一确定这个系统中某一个量，称之为 α, β 的积，并用记号 $\alpha\beta$ 表示．

上述三个运算具有以下的性质[1]：

1′) $\alpha+\beta=\beta+\alpha$,

2′) $(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$,

3′) 在系统 S 中存在一个零量 θ，具有性质

$$\alpha+\theta=\alpha,$$

4′) $a(\alpha+\beta)=a\alpha+a\beta$,

5′) $(a+b)\alpha=a\alpha+b\alpha$,

1) 希腊字母表示系统 S 中的任意量，而拉丁字母表示域 P 中的任意元素．

6′) $(ab)\alpha=a(b\alpha)$,

7′) $\theta\alpha=\theta$, $1\cdot\alpha=\alpha$, 此处1是域 P 中的单位元,

8′) 在 S 中存在这样的一组量 $\alpha_1, \alpha_2, \cdots, \alpha_n$, 使得 S 中每一个量均可唯一地表示成以下形式:

$$a_1\alpha_1+a_2\alpha_2+\cdots+a_n\alpha_n,$$

9′) $(a\alpha)\beta=\alpha(a\beta)=a(\alpha\beta)$,

10′) $\alpha(\beta+\gamma)=\alpha\beta+\alpha\gamma$, $(\beta+\gamma)\alpha=\beta\alpha+\gamma\alpha$.

在这个定义中, 任意域 P 的元素起着迄今为止实数所起的作用. 由条件 8′ 可知, 超复数系中的每一个元素被 P 中 n 个元素组所确定, 因之, 随着域 P 的选择, 超复数可定义为 n 个复数、n 个实数、n 个有理数等等.

前八个条件表明 S 组成域 P 上的有限维向量空间(参看第十六章 § 2), 此处 P 是代数的基本域.

条件 9′ 与 10′ 可以联合成下面等式的形状:

$$(a\beta+b\gamma)\alpha=a(\beta\alpha)+b(\gamma\alpha),$$
$$\alpha(a\beta+b\gamma)=a(\alpha\beta)+b(\alpha\gamma),$$

由这两个等式可知, 乘法的运算对于其每一因子都是线性的.

"超复数系"和"代数"这两个术语, 近年来偏重于应用第二个, 因为一般"超复数系"的元素就其性质来说与通常的数区别很大, 把它称之为"超复数"是不怎么合适的. "超复数系", "超复数"现在仅应用于最简单的代数, 例如对于普通的四元数.

由条件 1′—10′ 可见, 在代数中不假定乘法适合交换律及结合律, 也不假定存在单位元及可施行"除法".

在每一个代数 S 中都存在着基底, 即这样的元素组 $\alpha_1, \alpha_2, \cdots, \alpha_n$, 利用它, 代数中的所有元素都能唯一地表为系数属于基本域 P 的线性组合的形式 $a_1\alpha_1+a_2\alpha_2+\cdots+a_n\alpha_n$. 每一个代数可以有无限多个基底, 但每一个基底中元素的个数是同一的, 称这个数为代数的维数.

复数系看作实数体上的代数, 以 1 与 i 为其基底. 但数 2 与 $3i$, 1 与 $a+bi$(a, b 是任意实数, $b\neq0$)也可以当作基底.

设 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 是某个域 P 上代数的一组基底. 按照定义, 代数的每一元素均可唯一地表成

$$\alpha=a_1\varepsilon_1+a_2\varepsilon_2+\cdots+a_n\varepsilon_n$$

的形式. 设 $\beta=b_1\varepsilon_1+b_2\varepsilon_2+\cdots+b_n\varepsilon_n$ 是代数的另一个元素, 那么, 由于性质 1′—6′, 有

$$\alpha+\beta=(a_1+b_1)\varepsilon_1+(a_2+b_2)\varepsilon_2+\cdots+(a_n+b_n)\varepsilon_n.$$

类似地, 对于域 P 中任意的 a, 有

$$a\alpha=aa_1\varepsilon_1+aa_2\varepsilon_2+\cdots+aa_n\varepsilon_n.$$

因之, 代数中元素的加法以及域 P 中元素与代数的元素的乘法可以被上述公式完全决定. 代数中元素的相乘, 也不必需要知道任意两个元素如何乘法, 只要知道基底元素 ε_i 如何相乘即已足够. 事实上, 由于性质 9′ 及 10′,

$$(a_1\varepsilon_1+a_2\varepsilon_2+\cdots+a_n\varepsilon_n)(b_1\varepsilon_1+b_2\varepsilon_2+\cdots+b_n\varepsilon_n)=\Sigma a_ib_j\cdot\varepsilon_i\varepsilon_j.$$

乘积 $\varepsilon_i\varepsilon_j$ 中每一个都是代数中的元素, 因之, 可用基底元素表示出

$$\varepsilon_i\varepsilon_j=c_{ij1}\varepsilon_1+c_{ij2}\varepsilon_2+\cdots+c_{ijn}\varepsilon_n,$$

此处 c_{ijk} 表示代数的基本域 P 中的元素. 第一重足码表示头一个因子号数, 第二重足码表示第二个因子号数, 第三重足码表示这个元素用基底表示时系数的号数. 系数 c_{ijk} 叫做代数的构造常数, 因为它可以完全决定代数元素的运算.

很容易计算出 n 阶代数构造常数的个数. 每一个构造常数有三个号码 i, j, k. 因之, n 阶代数的构造常数的个数等于从 1, 2, $\cdots$, n 中每次取三个的排列总数, 故等于 n^3. 例如, 复数系在实数域上有基底 1, i. 由于等式

$$1\cdot1=1\cdot1+0\cdot i,\quad i\cdot1=0\cdot1+1\cdot i,$$
$$1\cdot i=0\cdot1+1\cdot i,\quad i\cdot i=-1\cdot1+0\cdot i,$$

故构造常数分别等于

$$c_{111}=1,\ c_{112}=0,\ c_{211}=0,\ c_{212}=1,$$
$$c_{121}=0,\ c_{122}=1,\ c_{221}=-1,\ c_{222}=0.$$

反之, 设给出某个域 P 的用三重足码表示的 n^3 个元素 c_{ijk} $(i, j, k=1, 2, \cdots, n)$, 那么可以把这些元素看作域 P 上代数的

构造常数，而且以等式 $e_ie_j=\sum_{k=1}^{n}c_{ijk}e_k$ 作为代数中乘法运算规则的定义.

上面我们已经看到，每一个代数通常有无限多个不同的基底.构造常数依赖于基底的选择，因此，不同的构造常数组可以确定同一个代数.

怎样的代数应该认为是不同的？怎样的应该认为是相同的？在代数论中，如果同一域 P 上的两个代数是同构的，则认为它们是相同的；也就是说，一个代数的元素与另一个代数的元素之间存在着这样的一一对应，使得头一个代数中的任意两个元素的和与积与第二个代数的相应元素的和与积对应，而域 P 中元素与头一个代数的元素的乘积与域 P 中的这个同一元素与第二个代数的相应元素的乘积相对应.

代数相同的这个定义表明，在代数理论里仅研究元素与元素组的可以借三种基本运算表出的那些性质. 换言之，代数理论研究对代数元素所施行的运算的性质，而不是研究代数元素的性质.

很容易证明，如果两个代数是同构的，那么，与一个代数的基底元素所对应的元素是另一代数的基底，而且，在相应基底下，构造常数也是相等的. 反之，如果同一域上的两个代数，在适宜的基底下有相同的构造常数，那么，这两个代数是同构的.

在所有代数中，结合代数直到今天仍然起着非常重要的作用. 所谓结合代数，就是乘法运算适合结合律 $\alpha(\beta\gamma)=(\alpha\beta)\gamma$ 的代数. 在本节中所叙述的正是这样的代数的性质. 在非结合代数中最有兴趣的是李代数，它的乘法满足下述性质：

$$\alpha\beta=-\beta\alpha,\ \alpha(\beta\gamma)+\beta(\gamma\alpha)+\gamma(\alpha\beta)=0.$$

我们所以对它有兴趣是由于它和李群间有密切的联系，而关于李群，我们在 §7 中已经谈论过.

矩阵代数 前面我们曾经指出，在超复数系理论发展的前期，人们的主要注意力集中在研究由于某种原因引起研究者兴趣的个别系统上. 这些系统中的某些，我们已经考查过. 大致在上一世

纪中叶，人们就开始研究在现今一般代数理论中起着主要作用的矩阵代数，这里，我们将简略地回忆一下矩阵运算的定义(参看第十六章§1).

域 P 上的一个矩阵是指域 P 中的元素 P 所排成长方形的一张表．两个矩阵说是相等的，如果其相应位置的元素相同．此处我们将仅讨论方阵，即行数与列数相同的矩阵．方阵的行数或者与它相等的列数叫做方阵的阶数.

为了相加两个阶数相同的方阵，应把它们的相应元素相加．用一个数去乘矩阵按定义等于这个数遍乘这个矩阵的所有元素．矩阵与矩阵相乘的定义比较复杂，两个 n 阶方阵的乘积是指这样的一个 n 阶方阵，其第 i 行第 j 列的元素等于头一个矩阵第 i 行的元素与第二个矩阵第 j 列的相应元素乘积的和．例如

$$\begin{pmatrix} a & b \\ a_1 & b_1 \end{pmatrix}\begin{pmatrix} x & y \\ x_1 & y_1 \end{pmatrix}=\begin{pmatrix} ax+bx_1 & ay+by_1 \\ a_1x+b_1y & a_1y+b_1y_1 \end{pmatrix}.$$

矩阵相乘的定义所以如此选择的原因，在第十六章中曾经叙述过.

按照上述定义，元素属于某个域 P 的全体 n 阶方阵所组成的量的系统，它们可以相加，可以与域中的元素相乘，它们彼此间也能相乘．经过不甚复杂的计算可以证明，代数定义中的性质 1′—10′，这里都满足．除此以外，容易证明矩阵乘法适合结合律．因此，元素属于给定域 P 的、阶数 n 给定的全体方阵组成这个域上的一个结合代数.

显然，等式

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}=a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}+b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}+c\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}+d\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

表明，等式右端的四个方阵组成二阶方阵代数的基底．一般说来，如果用符号 e_{ij} 表示第 i 行第 j 列位置的元素为 1 而其余位置的元素均为 0 的 n 阶方阵，那么，由于等式

$$\begin{pmatrix} a_{11}\cdots a_{1n} \\ \vdots \qquad \vdots \\ a_{n1}\cdots a_{nn} \end{pmatrix}=\sum_{i,j}a_{ij}e_{ij}$$

成立，故知方阵 ε_{ij} 组成 n 阶方阵代数的基底；由于方阵 ε_{ij} 的个数为 n^2，故 n 阶方阵代数的维数为 n^2. 基底矩阵 ε_{ij} 的乘法表有以下形式：

$$\varepsilon_{ij}\cdot\varepsilon_{jl}=\varepsilon_{il},\ \varepsilon_{ij}\cdot\varepsilon_{kl}=0,\ j\neq k,\quad i,\ j,\ k,\ l=1,\ 2,\ \cdots,\ n.$$

矩阵代数含有单位元，即单位方阵.

结合代数的表示 设对于域 P 上某个代数 A 的每一个元素唯一地确定同一域 P 上某个代数 B 的一个元素. 如果 A 中任意两个元素的和与积，都与 B 中相应的两个元素的和与积对应，而且域中的每一元素与代数 A 中元素的乘积，对应于域中的这个同一元素与代数 B 中相应元素的乘积，那么就说代数 A 同态地映象于代数 B 内. 结合代数 A 到 n 阶方阵代数里的一个同态映象叫做代数 A 的一个 n 次表示. 如果代数 A 的不同元素对应于方阵的不同元素，那么就说这个表示是真正的或者同构的. 当代数 A 用矩阵同构表示时，可以认为代数元素的运算可以归结为相应矩阵的运算. 因此，求代数表示的问题有重大意义. 这里我们仅讨论求代数表示的最简单的方法，然而，它在一般理论中有重要的作用.

给出某个代数 A，选择其一组基底 $\varepsilon_1,\ \varepsilon_2,\ \cdots,\ \varepsilon_n$，并假定 α 是 A 的任意元素. 乘积 $\varepsilon_1\alpha,\ \varepsilon_2\alpha,\ \cdots,\ \varepsilon_n\alpha$ 仍是 A 的元素，因之，应能用 $\varepsilon_1,\ \varepsilon_2,\ \cdots,\ \varepsilon_n$ 的线性表示. 设

$$\begin{aligned}
\varepsilon_1\alpha&=a_{11}\varepsilon_1+a_{12}\varepsilon_2+\cdots+a_{1n}\varepsilon_n,\\
\varepsilon_2\alpha&=a_{21}\varepsilon_1+a_{22}\varepsilon_2+\cdots+a_{2n}\varepsilon_n,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\varepsilon_n\alpha&=a_{n1}\varepsilon_1+a_{n2}\varepsilon_2+\cdots+a_{nn}\varepsilon_n.
\end{aligned}$$

这样，对于固定基底，每一个元素 α 可以对应到一个确定的矩阵 $\|a_{ij}\|$. 经过简单的计算，可以证明这个对应是代数 A 的一个表示. 这个表示常称做代数 A 的正则表示. 显然这个表示的次数等于代数的维数.

复数可以看作实数域上以 1, i 为基底的 2 维代数. 等式

$$1\cdot(a+bi)=a\cdot1+b\cdot i,$$

$$i\cdot(a+bi)=-b\cdot 1+a\cdot i$$

表明每一个复数 $a+bi$ 对应于矩阵 $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ 的对应是复数的正则表示. 类似地, 四元数的正则表示有以下形状:

$$a+bi+cj+dk\to\begin{pmatrix} a & b & c & d \\ -b & a & -d & c \\ -c & d & a & -b \\ -d & -c & b & a \end{pmatrix}.$$

复数和四元数的上述表示是真正的(同构的). 然而有例子表明, 正则表示并不永远是真正的. 但是, 如果代数含有单位元, 那么它的正则表示必然是真正的.

容易证明, 每一个结合代数都可包括在一个有单位元的代数中. 这个扩大代数的正则表示是真正的; 因之, 对于原来给出的代数来说, 这个正则表示也是真正的. 因此, 每一个结合代数都有用矩阵的真正表示.

上述求表示的方法对于建立代数的所有表示来说是不够的. 比较细致的方法是与代数的理想子代数的概念相联系着, 而理想子代数在近代数学上起着重要的作用.

代数的某些元素的集合 I 叫做一个右理想子代数, 如果它是一个线性空间, 而且用代数的任意元素右乘 I 中任意元素的结果仍然属于 I. 类似地(交换因子的顺序)可以定义左理想子代数. 同时是左、右理想子代数的叫做两边理想子代数. 显然, 代数中的零元素本身作成一个两边理想子代数, 称之为零理想子代数. 同样, 每一个代数本身也可看作其自己的理想子代数. 然而, 除了这两个当然的理想子代数外, 代数可能含有其他的理想子代数, 它的存在通常与代数的有趣的性质相联系.

设结合代数 A 含有右理想子代数 I. 在这个理想子代数中选择一组基底 $e_1, e_2, \cdots, e_m$. 因为在一般情形 I 仅是 A 的部分, 故 I 的基底元素的个数少于 A 的基底元素的个数. 设 α 是 A 中任意元素. 因为 I 是右理想子代数, 且 $e_1, e_2, \cdots, e_m$ 含于 I, 故 $e_1\alpha$,

$\varepsilon_2\alpha, \cdots, \varepsilon_m\alpha$ 也含于 I. 因之可以用基底 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_m$ 线性表示，即

$$\begin{aligned}\varepsilon_1\alpha &= a_{11}\varepsilon_1 + a_{12}\varepsilon_2 + \cdots + a_{1m}\varepsilon_m,\\ \varepsilon_2\alpha &= a_{21}\varepsilon_1 + a_{22}\varepsilon_2 + \cdots + a_{2m}\varepsilon_m,\\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \varepsilon_m\alpha &= a_{m1}\varepsilon_1 + a_{m2}\varepsilon_2 + \cdots + a_{mm}\varepsilon_m.\end{aligned}$$

命矩阵 $\|a_{ij}\|$ 与元素 α 对应，与前面一样，我们得到了代数 A 的一个表示. 这个表示的次数等于理想子代数 I 的基底元素的个数，因之，在一般情形将少于正则表示的次数. 显然，如果理想子代数本身是最小的，则利用这个理想子代数所得到的表示的次数也将是最小的. 由此可以看到，最小理想子代数在代数理论中起着特别重要的作用.

代数的构造 如上所述，每一个结合代数 A 可以用某阶的矩阵同构表示. 在这个表示中，与代数 A 的元素对应的全体矩阵本身组成一个代数，是给定阶数的矩阵代数的一部分. 如果一个代数的部分集合作成一个代数，则称之为原来代数的子代数. 因之，可以这样说，每一个结合代数都与矩阵代数的某个子代数同构.

虽然这个结果有原则性的益处，因为它把求所有代数的问题归结为求矩阵代数的所有子代数的问题，但是，它并不能对代数的构造问题给出直接的回答. 对于这个问题头一个一般的回答是上一世纪末由列夫斯克大学(塔尔吐斯克)教授摩林(1861—1941)于1900年在托姆斯克工艺大学给出的.

一个代数如果不含异于0及本身的两边理想子代数，那么称之为单纯代数. 摩林证明了复数域上维数≥2的单纯结合代数都与这个域上适当阶数的全矩阵代数同构.

继摩林之后，魏德邦在二十世纪初得到了一系列的结果，非常完善地阐明了任意域上代数的构造.

代数 A 的某些元素的集合(特别是代数 A 本身、它的某个理想子代数或子代数)叫做幂零的，如果存在这样一个自然数 s，使得这个集合中任意 s 个元素的乘积均等于零. 每一个结合代数都

有一个极大的幂零两边理想子代数，称之为代数的根．根等于零的代数叫做**半单纯的**．可以证明，每一个半单纯可以分解成特殊类型的单纯代数的和，这样，整个半单纯代数的研究可以归结为单纯代数的研究．最后，代数 A 叫做**可除代数**，如果在 A 中每一个形如 $ax=b(a\neq 0)$ 的方程都有解．

复数域上单纯代数的构造被上面提到的摩林定理完善描述出．如果基本域是任意的，那么有更一般的魏德邦定理成立：域 P 上维数大于等于 2 的每一个单纯代数都与元素属于 P 上某个可除代数的适当阶数的全矩阵代数同构．这样一来，魏德邦定理把求给定域 P 上单纯代数的问题归结为求域 P 上可除代数的问题．复数域上仅有一个可除代数，即复数域本身．由魏德邦定理可知，复数域上的每一单纯代数都与复数域上某个阶数的全矩阵代数同构，这就是摩林定理．

实数域上仅存在三个结合的可除代数：实数域本身、复数域、四元数代数．这个断言的证明不太简单，这里我们将不讨论它．由魏德邦定理可知，实数域上每一个单纯代数都与实数域上或者复数域上或者四元数代数上适当阶数的全矩阵代数同构．

由这些例子可以看出，摩林和魏德邦定理是如何阐述半单纯代数构造的．至于谈到有根的代数，所谓的魏德邦主要定理，对于它们有重大意义．按照这个定理，对基本域加上某些附加限制，在每一个有根 R 的代数 A 中，存在一个半单纯代数 L，使得 A 的每一个元素都可唯一地表为和 $\lambda+\rho(\lambda\in L,\ \rho\in R)$ 的形状，而且在某种意义下，子代数 L 在 A 中是唯一确定的．

上面叙述的一些主要定理，给出关于结合代数可能类型的完整观念，并且把它们的构造问题基本上归结为关于幂零代数构造的相应问题．后者的理论暂时还在形成过程中．

§13．李 代 数

在 §12 中我们曾说过，现在除结合代数的理论外，李代数的

理论也非常详细地被研究着．在李代数中，乘法满足以下的要求：

$$\alpha\beta=-\beta\alpha,\ \alpha(\beta\gamma)+\beta(\gamma\alpha)+\gamma(\alpha\beta)=0.$$

这样代数的重要性在于它们与李群(参考§7)有密切的联系，即与连续群的最重要类型有联系．正如我们前面看到的，李群在近世几何中起着重大的作用．按照李群和李代数理论的起源，我们对于所有实数以及所有复数作成的域上的李代数最感兴趣．

下面是李代数的一个最简单的例子．考虑给定阶数 n 的全部 n 阶方阵的集合．对于它们引入换位运算，即对于给定的矩阵 A, B, 命称之为换位子的 $AB-BA$ 与它们对应，并用符号 $[A, B]$ 表示 A, B 的换位子．

很容易证明

$$[A, B]=-[B, A]$$

$$[A, [B, C]]+[B, [C, A]]+[C, [A, B]]=0.$$

因之，给定阶数的全部方阵对于换位运算作成一个李代数．很明显，矩阵所组成的李代数的每一个子代数，即每一个矩阵的集合，对加法运算、乘以基本域中的数的乘法、换位运算均关闭，本身也是一个李代数．

关于对每一个抽象的李代数是否存在一个矩阵代数与其同构的问题，长时期来未得解决．直到1935年，著名代数学家切波塔廖夫的学生阿朵才给出了肯定的回答．

关于李群与李代数之间的对应关系，现在只谈其大概，而不涉及详细情形，也不给与严格的描述；并且仅限于李群与李代数均用矩阵表示的情形．

设 L 是某个矩阵的李代数．对于 L 中每一个矩阵 A, 令矩阵 $U=e^A=E+\frac{A}{1!}+\frac{A^2}{2!}+\cdots$ 与之对应．于是，用这样方法所得的全体矩阵的集合对于通常的矩阵乘法组成一个李群．反之，对于每一个李群，存在唯一的一个(就同构意义来说)李代数，使得与它对应的李群恰与所给李群同构．

为了简单起见，上面引出的关于李群和李代数之间对应关系

的定理不是精确的，而只是简化了的描述．实际上，关系 $U=e^A$ 仅对于与单位矩阵足够接近的 U 以及与零矩阵足够接近的 A 存在．严格的描述需要引进相当复杂的局部群与局部同构的概念．

因此，由李代数到其相应的李群的过渡，是借类似于方幂的运算来实现的，而由群到代数的过渡是借类似于求对数的运算来实现的．

若 L 是全体 n 阶方阵所作成的代数，那么与 L 对应的李群将是全体非退化方阵所作成的群，因为任意一个与单位方阵足够接近的 U 均可表成 $U=e^A$ 的形状．

矩阵 $A=\|a_{ij}\|$ 叫做斜对称的，如果它的元素之间适合关系式 $a_{ji}=-a_{ij}$．全体斜对称方阵组成一个李代数，因为若 A, B 是斜对称方阵，则 $AB-BA=[A, B]$ 以及 $\alpha A+\beta B$ 也将是斜对称方阵．

容易验证，对于每一个斜对称方阵 A，式子 e^A 将是一个正交方阵，而且对于每一个与单位方阵足够接近的正交方阵，均可表成上述指数的形式．因之，所有正交方阵所成群的李代数是全体斜对称方阵代数．

由解析几何知道，空间中围绕坐标原点的任意一个旋转都可借正交方阵给出，而且旋转的乘积与相应的矩阵的乘积相对应．换句话说，空间中围绕固定点的所有旋转所成群，与三阶正交方阵群同构．由此我们可得出结论：空间的旋转群的李代数是全体三阶斜对称方阵代数，亦即如下形状的方阵

$$A=\begin{pmatrix} 0 & -a & -b \\ a & 0 & -c \\ b & c & 0 \end{pmatrix}$$

所作成的李代数．

因为这样的每一方阵由三个数 a, b, c 完全确定，所以我们可以约定，用在坐标轴上投影为 a, b, c 的向量 $\boldsymbol{a}$ 表示．而且上述形状的方阵 A_1, A_2 的线性组合 $\alpha A_1+\beta A_2$ 显然与对应向量 $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ 的线性组合 $\alpha\boldsymbol{a}_1+\beta\boldsymbol{a}_2$ 相对应，而方阵 A_1, A_2 的换位子

$$[A_1,\ A_2]=A_1A_2-A_2A_1$$

$$=\begin{pmatrix}0 & -a_1 & -b_1\\ a_1 & 0 & -c_1\\ b_1 & c_1 & 0\end{pmatrix}\begin{pmatrix}0 & -a_2 & -b_2\\ a_2 & 0 & -c_2\\ b_2 & c_2 & 0\end{pmatrix}-\begin{pmatrix}0 & -a_2 & -b_2\\ a_2 & 0 & -c_2\\ b_2 & c_2 & 0\end{pmatrix}\begin{pmatrix}0 & -a_1 & -b_1\\ a_1 & 0 & -c_1\\ b_1 & c_1 & 0\end{pmatrix}$$

$$=\begin{pmatrix}0 & b_2c_1-b_1c_2 & a_1c_2-a_2c_1\\ b_1c_2-b_2c_1 & 0 & a_2b_1-a_1b_2\\ a_2c_1-a_1c_2 & a_1b_2-a_2b_1 & 0\end{pmatrix}$$

将与射影为 $b_1c_2-b_2c_1$, $a_2c_1-a_1c_2$, $a_1b_2-a_2b_1$ 的向量相对应，即与向量 $\boldsymbol{a}_1$ 与 $\boldsymbol{a}_2$ 的向量乘积相对应. 我们得到了一个著名的结果：通常的向量的集合对于向量加法、纯量与向量的乘法以及向量乘法组成一个李代数，与空间中围绕不动点的旋转群相对应. 这个事实再一次告诉我们，几何概念与空间的旋转群，换句话说，与刚体运动规律的联系是多么密切！

在上一世纪末以及本世纪初，对于李代数得到了许多与结合代数主要结果相类似的结果，虽然这里的证明和描述都是比较复杂的. 这样，经过李、基林、卡当等人的努力，在二十世纪初，对于李代数已经成功地建立起根和半单纯的概念，并且求出了实数体和复数体上所有单纯李代数. 到了三十年代初，卡当和魏尔建立了李代数的矩阵表示理论，是解决一系列问题的极好工具. 近十五年来，一些苏联数学家对李代数理论进行了研究，在这个领域中获得了不少极好的结果. 特别，他们对李代数表示的理论作出了重大的推进，彻底地解决了关于李代数的半单纯子代数的问题、关于建立具有给定根的代数问题等等.

§ 14. 环

在§ 11 中(310 页)我们曾经给出域的一般定义，即元素的任意一个集合，在其中定义了满足条件 1—10 的加法与乘法的运算. 在这个定义中，如果去掉关于商的存在的条件 10 以及乘法适合交换律和结合律的条件 7 和 8，那么我们便得到近世代数学中最重

要概念之一的环的概念的定义.

每一个域以及每一个代数,如果仅考虑其加法与乘法的运算,是一个环. 环的最简单的例子是全体有理整数集对于通常的加法与乘法的运算. 所有形如 $a+bi$, $a+b\sqrt{2}$, $a+b\sqrt[3]{2}+c\sqrt[3]{4}$ 等等的数集对于通常的加法与乘法也都作成环,此处 a, b, c 是任意有理整数. 由于这些环的元素是数,故称之为数环. 这些环的某些重要性质及其应用在第四章(第一卷)及第十章(第二卷)曾经讨论过.

然而,也存在着非数环的重要的类. 例如,给定变量 x_1, x_2, …, x_n 系数属于某个固定的环或域的全体多项式的集合,定义在某区域上全体连续函数的集合、线性空间或者希尔伯特空间的线性变换的集合,对于通常定义的它们的加法与乘法运算都作成环.

数环的算术性质是介乎代数本身与数论本身之间的深刻的代数数论的研究对象. 多项式环性质的研究被称之为多项式理想子环论,与解析几何的高等部分有着密切的联系. 最后,函数环与变换环在泛函分析上起着基本作用(参看第十九章).

以这些及其他一些具体理论为基础,近百年来,一般环的理论以及拓扑环的理论发展非常迅速.

由于篇幅的限制,下面我们仅指出环论中的某些个别结果.

理想子环 某个环 K (不一定是结合的)的元素的部分集合 I 叫做 K 的理想子环,如果 I 中任意两个元素的差以及 I 中任意元素 x 与 K 中任意元素 a 的乘积 ax, xa 均属于 I.

每一个理想子环是环的这样一个部分,它本身对于在环中给出的加法与乘法的运算做成一个环. 这样的部分叫做已知环的子环,也就是说,每一个理想子环同时也是子环. 其逆通常是不成立的.

一个环的任意多个理想子环的交集仍是这个环的一个理想子环. 特别,环中含有某个固定元素 a 的所有理想子环的交集仍是这个环的理想子环. 这个理想子环叫做元素 a 生成的主理想子环,并用符号(a)表之.

两个或者若干个元素生成的理想子环的概念，可用与上面同样的方式定义．很容易证明，如果可换的结合环有单位元素，那么由元素 $a_1, a_2, \cdots, a_n$ 所生成的理想子环将简单地是环中可以写成和 $x_1a_1+x_2a_2+\cdots+x_na_n$ 形状的所有元素所组成，此处 $x_1, x_2, \cdots, x_n$ 是环的任意元素．特别，在有单位元的可换结合环中，主理想子环(a)是 a 的倍数的所有元素，亦即 xa 形状的所有元素所组成．

在所有有理整数所做成的环中，每一个理想子环都是主理想子环．在系数属于某个域的一个未知量的多项式环中、在一切形如 $a+bi$(此处 a, b 是有理数)的复数环中以及在其他一些环中也具有这样的性质．然而，在系数为有理数的未知量 x, y 的多项式环中，没有常数项的两个未知量 x, y 的所有多项式的集合，已经不是主理想子环．

与在群论中对不变子群所做过的类似，对于环 K 的每一个理想子环 I，可以建立商环 K/I．这可用下面方法来做．环 K 中元素 a, b 叫做关于理想子环 I 等价，如果它们的差 $a-b$ 属于 I，这时用符号 $a\equiv b(I)$表示．很容易验证，这个等价的关系是自反的、对称的、传递的(参看第十五章)，因之，K 的所有元素可以按关于理想子环 I 等价关系分成类．现在，让我们来考虑以这些类作为元素的新的集合，对于它引进和与乘积的概念：所谓两个类的“和”是指这样一个类，它含有分别属于这两个类的某两个元素的和，而两个类的乘积是指这样一个类，它含有分别属于这两个类的某两个元素的乘积．由理想子环的定义可知，和与乘积的这样定义，事实上是与代表选择无关的，并且最后所有类的集合作成一环．

在环论中，商环的作用完全类似于在群论中商群的作用．特别，由已知环建立商环，是形成具有各种不同性质的环的一种特别方便的方法．此外，例如很容易证明，任意一个可换环都能与以有理整数为系数的适当个数未知量的多项式环的商环同构．

环的算术性质 在数环以及在域中，若干个元素的乘积，当且仅当其中有一个因子等于零时才能等于零. 在任意环中，这个事实可以不成立，例如，两个不等于零的矩阵的乘积可以是零. 如果在某个环中，$a\neq0$, $b\neq0$. 但是有 $ab=0$, 那么就说 a 与 b 是零因子. 如果在一个环中没有这样性质的元素，就说这个环是没有零因子的环.

在环中研究可除性时，通常假定这个环是可换的，而且没有零因子. 这样的环照例叫做整域. 上面提到的数环以及多项式环都是整域.

设 K 是某个整域. 我们说，元素 a 在 K 中能被 b 整除，如果 $a=bq$, $q\in K$. 由此直接可知，被 b 整除的若干个元素的和仍能被 b 整除，K 中若干个元素的乘积中如果有一个因子能被 b 整除，则这个乘积也能被 b 整除. 在环论中引入类似于第十章(第二卷)所讲过的素数概念的素元概念，通常是比较复杂的. 这时必须先引入相伴元的概念. 环中的元素 a, b 叫做相伴的，如果 a 被 b 整除，b 也被 a 整除. 命 $a=bp_1$, $b=aq_2$, 则有 $ab=ab\cdot q_1q_2$, 亦即 $q_1q_2=e$, 此处 e 是整域 K 的单位元. 因此，称相伴元的商为单位元因子. 整域中的每一个元均能被任意一个单位元因子整除. 在有理整数环中，单位元因子是 ±1, 在一切 $a+bi$(a, b 是整数)形状的元素所组成的环中，单位元因子是 ±1, $\pm i$.

整域 K 中的每一个元素均有形如 $a=a\varepsilon\cdot\varepsilon^{-1}$ 的分解，此处 ε 是某个单位元因子. 这样的分解叫做显然的分解. 如果 K 中的元素 a 没有其他分解，那么，就称 a 为 K 的素元，或者叫作 K 的不可分解元. 由于整数分解为素因子的唯一性定理具有重要意义，于是找这样的环的类(其中也包括非可换环)，使得类似定理在其中也成立，是有意义的. 例如，在主理想子环中，即其中每一个理想子环都是主理想子环的整域，这个定理成立.

理想子环这一概念本身的出现，是与分解为素因子唯一性问题相联系的. 大约在上一世纪中叶，德国数学家库莫尔企图证明费尔马的关于方程 $x^n+y^n=z^n$ 当 $n\geqslant3$ 时没有非零的整数解的著

名命题时，考虑了形如 $a_0+a_1\xi+\cdots+a_n\xi^{n-1}$ 的数，此处 $\xi=\cos\frac{2\pi}{n}+i\sin\frac{2\pi}{n}$ 是方程 $x^n=1$ 的解，而 $a_0, a_1, \cdots, a_n$ 是普通整数．这样形状的数组成一个整域．库莫尔最初把这个整域中分解为素因子的唯一性定理看作是一个显然的命题．在这个基础上库莫尔证明了费尔马命题．然而，在进一步检查时，发现了上述分解为素因子的唯一性定理不成立．为了希望保持素因子分解的唯一性定理成立，库莫尔被迫讨论整域中数的分解，而这分解中的因子不在该整域中．他称这些数为理想数．后来在建立一般理论时，引进了被某个理想数整除的、整域的元素的集合以代替理想数．

发现在数环中分解为素因子的唯一性不成立，是上一世纪所发现的最有兴趣的事实之一，并由此产生了丰富的代数数论．

这个理论的极好应用之一是在第十章(第二卷)末指出的对分解普通整数为平方和问题的应用．苏联数学家佐洛塔辽夫、沃洛诺依、维诺格拉朵夫以及切波塔廖夫的工作在数环的理论发展中起着非常重要的作用．

代数流形 理想子环论的另一个发源地是代数几何．当我们第一次接触二次曲线的理论时，对下面事实常常会感到惊奇：一个双曲线是指互不相连的两个曲线的全体——每一个叫做双曲线的一支，同时两条直线叫做一个分解的二次曲线．术语上的区别在代数中找到解释：如果把曲线的方程表为 $f(x, y)=0$ 的形状，此处 $f(x, y)$ 是 x, y 的多项式，那么，在第一种情形，这个方程的左端是二次既约多项式；而在第二种情形，左端是两个一次因式的乘积．一个曲线叫做既约的，如果它的方程可借既约多项式 $f(x, y)$ 表出，相反的情形，曲线叫可约的．

同样问题对空间曲线进行讨论时，就变得较为复杂．空间曲线可以用含有两个方程的方程组 $f(x, y, z)=0$, $g(x, y, z)=0$ 表示，这时，f, g 远不是唯一地被曲线所决定．这里，自然发生这样的问题，应该把什么样的曲线叫做既约曲线？

理想子环论给出自然的回答．设 $f_1, f_2, \cdots$ 是变量为 x, y, z 的复系数多项式的某个集合，使得这个集合中每一个多项式都等于零的点 (x, y, z) 的集合(复空间的点)叫做给定的这些多项式所确定的一个代数流形．用 M 表示这个流形，并讨论被 M 中任何点代入都等于零的一切变量 x, y, z 的多项式集合．很容易看出，所有这样多项式的集合 I 是变量 x, y, z 的多项式环的一个理想子环．此外，这个理想子环还具有这样性质：如果某个多项式的方幂属于 I，那么，这个多项式本身也将属于 I，可是，不同的多项式的集合，可以确定同一的代数流形，然而，代数流形与具有上面提到的补充性质的理想子环之间的对应是一一对应．

因此，在研究流形的性质时，很自然的不讨论它们的"方程"，而讨论其对应的理想子环．如果理想子环 I 可以表成某两个理想子环 I_1, I_2 的交集，那么，流形 M 将可表成与 I_1, I_2 对应的流形 M_1, M_2 的并集．由此可以看出，流形 M 自然地叫做既约的，如果与它对应的理想子环 I 不能表成包含 I 的两个不同理想子环的交集．曲线分解为阶数较低的曲线与流形分解为既约流形，它们现在将与相应的理想子环表为不可分解的理想子环的交的形式相对应．这样分解的唯一性和可能性的问题是代数流形以及一般的理想子环论的最初问题之一．

非可换环的构造　每一个代数，同时也是对于它的加法与乘法运算的一个环．因此，代数论的许多基本概念以及结果对任意环也成立．然而，代数论的一些比较细致的结果，特别是类似于摩林-魏德邦定理(参看 § 11)转移到任意环上来，有很多困难．最近10—15年来，才部分地克服了一些．这里所指的困难首先是在于寻求环的根的定义，使得根为零的环与半单纯代数有某种类似，并且在任何情形，可以由环论的定理作为其特殊情形得出相应的代数论的结果．现在，在环论上已经有很多根的定义，利用它，在某种限制下建立起半单纯环的构造论．正如已经指出的，对非可换环论的兴趣在很大程度上受到作用子环论在泛函分析中非常显著的意义所刺激．

§15. 格

正如读者已经知道的，元素的集合叫做偏序集，如果对于它的某些元素对，定义了一元素在另一元数之前，或者说从属于另一元素，并且适合以下条件：1）每一个元素从属于其本身；2）由 a 从属于元素 b 以及 b 从属于元素 a，可得出 a 恒等于 b；3）由 a 从属于 b 以及 b 从属于 c 可得出 a 从属于 c. 从属关系通常用符号 $\leqslant$ 表示. 某个集合的所有子集的集合是偏序集的重要例子，此处，从属关系是一个子集是另一个子集的一部分.

如果从属关系对于偏序集的每一对元素均被定义，那么，称这个集合为有序（线性）集. 例如，实数集是有序集，此处关系 $a\leqslant b$ 表示 a 不大于 b. 相反的，含有多于一个元素的某个集合的所有子集的集合所做成的偏序集不是有序集，因为没有共同元素的两个子集是不能比较的.

设某个偏序集 M 具有以下性质：对 M 的每一对元素 a, b, 有唯一的最小公共大元 c, 即 $a\leqslant c$, $b\leqslant c$, 并且对 M 中任意元素 d, 只要 $a\leqslant d$, $b\leqslant d$, 就有 $c\leqslant d$. 这时称 M 为上半格，而称元素 c 为 a, b 的"和". 很易证明，这个"加法"具有下面的性质：

$$a+b=b+a,\ (a+b)+c=a+(b+c),\ a+a=a. \tag{13}$$

非常妙的是可以证明其反面. 设在某个集合中定义了具有性质(13)的加法运算，那么，当 $a+b=b$ 时，称元素 a 从属于 b. 这样，我们便得出一个偏序集，而且在这个偏序集中，元素 a, b 的唯一最小公共大元将是 $a+b$.

类似地，可以定义下半格. 这时，讨论最大公共小元以代替最小公共大元，并且称 a, b 的最大公共小元为 a, b 的"积"，这个运算也具有上面"加法"的性质，即

$$ab=ba,\ (ab)c=a(bc),\ aa=a. \tag{14}$$

既是上半格又是下半格的一个偏序集叫做一个格. 按照前面所说过的，在每一个格里，可以定义满足条件(13), (14)的两种运

算．然而，这两种运算彼此间是有联系的，因为关系 $a\leqslant b$ 在格中可以写成 $a+b=b$，$ab=a$ 的任一形式．换言之，在格中等式 $a+b=b$ 与 $ab=a$ 应该是等价的，也就是说，上面条件可以写成代数形式的等式

$$a+ab=a,\ a(a+b)=a. \tag{15}$$

由于这个缘故，格的研究便成为研究具有适合条件(13)，(14)，(15)的两个运算系统的纯粹代数问题．用代数的方法研究格论的意义，粗略的说，在于：在个别情形的某种具体的格的特点得以表成元素间某种代数关系，因而得以利用群论和环论上的丰富工具．

上面已经提到过，某个集合的所有子集的集合是一个偏序集．不难证明，它将是一个格，如果规定格的和是子集的并集，而格的积是子集的交集．如果不是讨论全部而仅讨论某些子集，那么，可以得出各式各样的格．例如，任意一个群的所有子群的集合是一个格，所有不变子群的集合也是一个格，任意一个环的所有子环的集合，所有理想子环的集合等等都是格．特别是，在一个群的所有不变子群所成格中，以及在一个环的所有理想子环所成格中，除了基本等式(13)，(14)，(15)成立外，下面的所谓模律

$$a(ab+c)=ab+ac$$

也成立．具有模律的格(戴德金格)的理论在一般格论中占有重要位置．

群论与环论中很大一部分定理是陈述子群、不变子群、理想子环的分布情况的．因之，这些定理可以作为子群或者理想子环的格的定理重新叙述．在某些限制下，类似的定理对于一般的格也成立．这种方法，能够把群论、环论以及其他科目中的某些重要定理转移到格论中来．另一方面，利用格论的工具，反过来在群论和环论中寻求具体的格的性质，也是很有益的．

格论的出现不是太久的事，在本世纪二十年代到三十年代还不象群论那样有重要的应用．然而，现在格论已经形成数学的一个重要分支，并且是一个有巨大内容和重要问题的分支．

§16. 一般代数系统

在前面几节曾经企图给出这样的概念：把代数方法应用于问题的更广范围，是如何引导出被代数所研究的元素的系统的扩展，也引导出代数运算概念本身的推广．在这方面，被罗巴切夫斯基关于几何基础的著作以及一般集合论的发展所刺激的公理方法的发展，起着巨大的作用．

这方面的基本成果之一是逐渐地总结出代数运算、代数系统的一般概念，并且累积了关于定义代数系统方面的一些重要事实．代替在中学所定义的具体的代数运算（这个定义很大一部分是关于数的），在近世代数中是从运算的一般概念出发．即给出元素的某个系统 S，并且给出一个规则，对于 S 中取一定顺序的 m 个元素 $a_1, a_2, \cdots, a_m$，在 S 中确定唯一的一个元素 a．这时，就说在 S 中给出了一个 m 元运算，并说元素 a 是施行这个运算于元素 $a_1, a_2, \cdots, a_m$ 上的结果．定义了一个或者若干个运算的元素的集合，叫做一个代数系统．代数的主要问题之一是研究代数系统并对之进行分类．然而这样形式的问题具有太一般的性质．事实上，在现在真正重要的并且具有丰富内容的仅是某些特殊的代数系统．例如，具有一个运算的代数系统，暂时只有在本章的§§1—10中介绍过的群论在数学科学中根深蒂固地成长起来，而在具有两个或者多于两个运算的代数系统中，域、代数、环以及格的理论有着重要意义．然而，由于这种或者那种原因，实际被讨论的代数系统的数目不断增多．同时，代数的某些典型的部分，例如，关于同态的理论、关于自由系统以及自由并的理论、关于直并的理论以及近来关于根的理论，也都转到一般的代数系统中来．这就可以作为代数的一个新的部分来谈论这些理论．

在研究整个代数科学的特点时，常常作为其特殊性质，着重指出其缺乏连续性的概念．因之承认代数科学主要是关于离散性的．这种观点，无疑地反映出代数的重要客观性质之一．在现实

世界中，连续性与不连续性是辩证的统一．但是，为了认识现实，有时需要把它分成几部分，然后分别地研究这些部分．因之，代数单方面的注意到离散关系，不能够认为这是它的缺点．

从群论的例子中可以看到，个别的代数科目不仅给出计算技巧的工具，而且也给出了表达深刻的自然界规律的语言．然而，代数各个部门除了对于物理、化学、结晶学以及其他一些科学的直接实践意义外，在数学本身，代数也占有重要地位．用杰出的苏联代数学家切波塔廖夫的话来说，代数是数学中发生的许多新的思想和概念的摇篮，它显著地丰富了并发展了数学的许多部门，这些部门已成为物理与技术科学的共同基础．

文　　献[1)]

П. С. Александров, Введение в теорию групп. Учпедгиз, 1952.

Ван дер Варден, Современная алгебра, ч. I и II. Гостехиздат, 1947.

Н. Джекобсон, Теория колец. ИЛ, 1947.

А. Г. Курош, Теория групп. Гостехиздат, 1953.

Л. С. Понтрягин, Непрерывные группы. Гостехиздат, 1954.

在苏联大百科全书"代数"、"群论"、"数学"等篇文章中，包括了简短的历史知识以及基本方法问题的讨论．苏联数学家在代数领域中的成就，叙述到1947年为止，刊登在"三十年来的苏联数学"文集中．Гостехиздат, 1948.

有关费得洛夫群的文献已在280页上指出．

刘绍学　吴品三　王隽骧　译
沈信耀　李培信　江嘉禾

秦元勋　校

刘绍学　复校

1) 代[illegible]一般教程在第四章及第十六章的文献中已经指出．